JOURNAL OF CONSULTING PSYCHOLOGY

Edited By
Laurance F. Shaffer
Teachers College, Columbia University

Advisory Editors

Edgar A. Doll
Bellingham, Washington

Ann Magaret Garner
University of Illinois

William A. Hunt
Northwestern University

E. Lowell Kelly
University of Michigan

Morris Krugman
Board of Education
New York, N. Y.

Bertha M. Luckey
Board of Education
Cleveland, Ohio

Fred McKinney
University of Missouri

Anne Roe
New York, N. Y.

Carl R. Rogers
University of Chicago

R. Nevitt Sanford
University of California

Edward Joseph Shoben, Jr.
Teachers College
Columbia University

J. R. Wittenborn
Rutgers University

Managing Editor
Lorraine Bouthilet

Bureau of Edu. & Psyl. Res
6 · 3 · 81
Acc. J692

VOLUME 19

1955

Published Bimonthly By

The American Psychological Association, Inc.
1333 Sixteenth St., N. W.
Washington 6, D. C.

Printed At
Lancaster, Pennsylvania

$7.50

Copyright © 1955 by the American Psychological Association, Inc.

Contents of Volume 19

Altus, Grace T. Relationships Between Verbal and Nonverbal Parts of the CTMM and WISC - 143

Armitage, S. G., Greenberg, P. D., Pearl, D., Berger, D. G., and Daston, P. G. Predicting Intelligence from the Rorschach - - - - - - - - - - - - - - - - - - - 321

Arnhoff, F. N. See *Hunt, W. A.*

Auld, F., Jr. See *Eron, L. D.*

Ax, A. F., and Colley, W. H. Temporal Acuity of Vision, Audition, and Touch in Psychogenic and Neurogenic Pathology - 455

Ax, A. F. See *Stauffacher, J. C.*

Baer, M. See *Buss, A. H.*

Baker, C. T., Sontag, L. W., and Nelson, Virginia L. Specific Ability in IQ Change - - - 307

Baker, R. W. See *Scherer, I. W.*

Barron, F. Threshold for the Perception of Human Movement in Inkblots - - - - - - 33

Barron, F., and Leary, T. F. Changes in Psychoneurotic Patients With and Without Psytherapy - 239

Barry, J. R., Sells, S. B., and Trites, D. K. Psychiatric Screening of Flying Personnel with the Cornell Word Form - 32

Becker, W. C., and McFarland, R. L. A Lobotomy Prognosis Scale - - - - - - - - - 157

Bendig, A. W. The Reliability of Adjustment Ratings and the Length of Case Histories - 463

Bendig, A. W., and Hamlin, R. M. The Psychiatric Validity of an Inverted Factor Analysis of Rorschach Scoring Categories - - - - - - - - - - - - - - - - - - 183

Bendig, A. W. See *Klugh, H. E.*

Berger, D. G. See *Armitage, S. G.*

Bernstein, L., and Chase, P. H. The Discriminative Ability of the Blacky Pictures with Ulcer Patients - 377

Bieri, J., Blacharsky, E., and Reid, J. W. Predictive Behavior and Personal Adjustment - 351

Bingham, J. L. See *Calvin, A. D.*

Blacharsky, E. See *Bieri, J.*

Bolin, B. J. A Comparison of Raven's Progressive Matrices (1938) with the ACE Psychological Examination and the Otis Gamma Mental Ability Test - - - - - - - - - 400

Borgatta, E. F. See *Eschenbach, A. E.*

Bordin, E. S. Ambiguity as a Therapeutic Variable - - - - - - - - - - - - - - - 9

Bordin, E. S. See *Harway, N. I.*

Boyd, Nancy A., and Mandler, G. Children's Responses to Human and Animal Stories and Pictures - 367

Broen, W. E., Jr. Personality Correlates of Certain Religious Attitudes - - - - - - - 64

Buss, A. H., Wiener, M., Durkee, Ann, and Baer, M. The Measurement of Anxiety in Clinical Situations - 125

Calvin, A. D., Koons, P. B., Jr., Bingham, J. L., and Fink, H. H. A Further Investigation of the Relationship Between Manifest Anxiety and Intelligence - - - - - - - - 280

Calvin, A. D. See *Schulz, R. E.*

Cartwright, D. S. Success in Psychotherapy as a Function of Certain Actuarial Variables - 357

Charen, S. The Awareness of Hostile Feelings in Patients by Their Nurses - - - - - - 290

Chase, P. H. See *Bernstein, L.*

Churchill, Ruth, and Crandall, V. J. The Reliability and Validity of the Rotter Incomplete Sentences Test - 345

Cohen, J. The Efficacy of Diagnostic Pattern Analysis with the Wechsler-Bellevue - - - 303

Colley, W. H. See *Ax, A. F.*

Corsini, R. J., Severson, W. E., Tunney, T. E., and Uehling, H. F. The Separation Capacity of the Rorschach - 194

Crandall, V. J. See *Churchill, Ruth*

Dahlstrom, W. G. See *Goodstein, L. D.*

Darling, C. D See *Summerskill, J.*

Daston, P. G. See *Armitage, S. G.*

Davids, A. Alienation, Social Apperception, and Ego Structure - - - - - - - - - 21

Davids, A. Relations Among Several Objective Measures of Anxiety Under Different Conditions of Motivation - 275

Davids, A., and Eriksen, C. W. The Relation of Manifest Anxiety to Association Productivity and Intellectual Attainment - - - - - - - - - - - - - - - - - 219

Davids, A., Henry, A. F., McArthur, C. C., and McNamara, L. F. Projection, Self Evaluation, and Clinical Evaluation of Aggression - - - - - - - - - - - - - - 437

Deabler, H. L. See *Price, A. C.*

Desai, M. M. The Relationship of the Wechsler-Bellevue Verbal Scale and the Progressive Matrices Test - 60

Dittmann, A. T. See *Harway, N. I.*

du Mas, F. M. Clinical Statements as Scientific Propositions and Social Decisions - - - 255

Durkee, Ann. See *Buss, A. H.*

Dymond, Rosalind F. Adjustment Changes in the Absence of Psychotherapy - - - - 103

Edwards, A. L. Social Desirability and Q Sorts - - - - - - - - - - - - - - 462

Eriksen, C. W. See *Davids, A.*

Eron, L. D., Sultan, Florence, and Auld, F., Jr. The Application of a Psychometric Scoring Procedure to a Group Modification of the Thematic Apperception Test (N-TAT) - - 83

Eschenbach, A. E., and Borgatta, E. F. Testing Behavior Hypotheses with the Rorschach: An Exploration in Validation - - - - - - - - - - - - - - - - - - 267

Estes, Betsy Worth. Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children - 225

Ewert, Josephine C. See *Green, M. W.*

Fairweather, G. W. See *Moran, L. J.*

Filmer-Bennett, G. The Rorschach as a Means of Predicting Treatment Outcome - - - 331

Fink, H. H. See *Calvin, A. D.*

Finney, B. C. The Diagnostic Discrimination of the "Basic Rorschach Score" - - - - - 96

Firestone, R. W. See *Harris, D. H.*

Flavell, J. H. Repression and the "Return of the Repressed" - - - - - - - - - - 441

Folsom, Angela T. See *Pruyser, P. W.*

Foundeur, M. R. See *Kugelmass, S.*

Gibby, R. G., Stotsky, B. A., Harrington, R. L., and Thomas, R. W. Rorschach Determinant Shift Among Hallucinatory and Delusional Patients - - - - - - - - - - - 44

Giedt, F. H. Comparison of Visual, Content, and Auditory Cues in Interviewing - - - - 407

Gleason, W. J. Direction of Perceived Movement in Males and Females - - - - - - 8

Gluck, M. R. Rorschach Content and Hostile Behavior - - - - - - - - - - - - 475

Goldberger, L. See *Goodstein, L. D.*

Goodstein, L. D., and Goldberger, L. Manifest Anxiety and Rorschach Performance in a Chronic Patient Population - - - - - - - - - - - - - - - - - - - 339

Goodstein, L. D., Spielberger, C. D., Williams, J. E., and Dahlstrom, W. G. The Effects of Serial Position and Design Difficulty on Recall of the Bender-Gestalt Test Designs 230

Green, M. W., and Ewert, Josephine C. Normative Data on Progressive Matrices (1947) - 139

Greenberg, P. D. See *Armitage, S. G.*

Guze, S. B. See *Saslow, G.*

Halpern, H. M. Empathy, Similarity, and Self-Satisfaction - - - - - - - - - - 449

Hamlin, R. M. See *Bendig, A. W.; Powers, W. T.*
Hardyck, C. D. See *Sarbin, T. R.*
Harrington, R. L. See *Stotsky, B. A.*
Harris, D. H., Firestone, R. W., and Wagner, C. M. Brief Psychotherapy and Enuresis - - 246
Harris, Pearl. Validity of the Grassi-Fairfield Block Substitution Test in Differential Diagnosis - 330
Harway, N. I., Dittmann, A. T., Raush, H. L., Bordin, E. S., and Rigler, D. The Measurement of Depth of Interpretation - - - - - - - - - - - - - - - - - - 247
Henry, A. F. See *Davids, A.*
Hilliard, Jessamine. See *Marsh, J. T.*
Hirschstein, R., and Rabin, A. I. Reactions to Rorschach Cards IV and VII as a Function of Parental Availability in Childhood - - - - - - - - - - - - - - - - 473
Holsopple, J. Q. See *Lorr, M.*
Howell, R. J. Changes in Wechsler Subtest Scores with Age - - - - - - - - - - - 47
Hunt, W. A., and Arnhoff, F. N. Some Standardized Scales for Disorganization in Schizophrenic Thinking - 171
Hurley, J. R. The Iowa Picture Interpretation Test: A Multiple-Choice Variation of the TAT - 372
Jenkins, R. L. See *Lorr, M.*
Johnson, Laverne C., and Stern, J. A. Rigidity on the Rorschach and Response to Intermittent Photic Stimulation - 311
Jourard, S. M., and Remy, R. M. Perceived Parental Attitudes, the Self, and Security - - 364
Kahn, T. C. Cross Validation of the Organic Brain Pathology Scale for a Test of Symbol Arrangement - 130
Keller, J. E. See *Rosenblum, S.*
Klugh, H. E., and Bendig, A. W. The Manifest Anxiety and ACE Scales and College Achievement - 487
Koons, P. B., Jr. See *Calvin, A. D.*
Kostlan, A. A Reply to Patterson - 486
Kubly, Delores. See *Martin, B.*
Kugelmass, S., and Foundeur, M. R. Zaslow's Test of Concept Formation: Reliability and Validity - 227
Lawrence, J. F. See *Stotsky, B. A.*
Leary, T. F. See *Barron, F.*
Levitt, E E., and Lyle, W. H., Jr. Evidence for the Validity of the Children's Form of the Picture-Frustration Study - 381
Levy, L. H. Movement as a "Rhetorical Embellishment" of Human Percepts - - - - - 469
Liechti, R. See *Marsh, J. T.*
Loos, F. M. See *Newman, J. R.*
Lorr, M., Holsopple, J. Q., Jenkins, R. L., and O'Connor, J. P. Factors of Change in Lobotomized Chronic Schizophrenic Patients - - - - - - - - - - - - - - 39
Lyle, W. H., Jr. See *Levitt, E. E.*
McArthur, C. C. See *Davids, A*
McFarland, R. L. See *Becker, W. C.*
McGaughran, L. S. See *Moran, L. J.*
McGowan, C. See *Martin, B.*
McNamara, L. F. See *Davids, A.*
Mandler, G. See *Boyd, Nancy A.*
Mark, J. C. See *Morrow, R. S.*
Marks, J. B. See *Stauffacher, J. C.*
Marsh, J. T., Hilliard, Jessamine, and Liechti, R. A Sexual Deviation Scale for the MMPI 55

Martin, B., and Kubly, Delores. Results of Treatment of Enuresis by a Conditioned Response Method - - - - - - 71
Martin, B., and McGowan, B. Some Evidence on the Validity of the Sarason Test Anxiety Scale - - - - - - 468
Matarazzo, J. D. MMPI Validity Scores as a Function of Increasing Levels of Anxiety - 213
Matarazzo, J. D., and Phillips, Jeanne S. Digit Symbol Performance as a Function of Increasing Levels of Anxiety - - - - - - 131
Matarazzo, J. D. See *Saslow, G.*
Matarazzo, Ruth G. The Relationship of Manifest Anxiety to Wechsler-Bellevue Subtest Performance - - - - - - 218
Mayzner, M. S., Jr., Sersen, E., and Tressell, M. E. The Taylor Manifest Anxiety Scale and Intelligence - - - - - - 401
Meltzer, M. L. See *Vernier, Claire M.*
Meyer, M. M., and Tolman, Ruth S. Correspondence Between Attitudes and Images of Parental Figures in TAT Stories and in Therapeutic Interviews - - - - - - 79
Meyer, M. M., and Tolman, Ruth S. Parental Figures in Sentence Completion Test, in TAT, and in Therapeutic Interviews - - - - - - 170
Mill, C. R., and Turner, C. J. The Measurement of Primary Mental Abilities by the Columbia Mental Maturity Scale - - - - - - 472
Miller, J. S., and Scodel, A. The Diagnostic Significance of Usual and Unusual TAT Stories - - - - - - 91
Moran, L. J., Fairweather, G. W., Morton, R. B., and McGaughran, L. S. The Use of Demographic Characteristics in Predicting Response to Hospitalization for Tuberculosis - - - - - - 65
Morrow, R. S., and Mark, J. C. The Correlation of Intelligence and Neurological Findings on Twenty-Two Patients Autopsied for Brain Damage - - - - - - 283
Morton, R. B. See *Fairweather, G. W.*
Mosak, H. H. Language and the Interpretation of "Sexual" Symbolism - - - - - - 108
Mussen, P. H., and Scodel, A. The Effects of Sexual Stimulation Under Varying Conditions on TAT Sexual Responsiveness - - - - - - 90
Nelson, Virginia L. See *Baker, C. T.*
Newman, J. R., and Loos, F. M. Differences Between Verbal and Performance IQ's with Mentally Defective Children on the Wechsler Intelligence Scale for Children - - - 16
Nunnally, J. C. A Systematic Approach to the Construction of Hypotheses About the Process of Psychotherapy - - - - - - 17
O'Connor, J. P. See *Lorr, M.*
Papania, N. See *Rosenblum, S.*
Parsons, E. T. Relationships Between the Rosenzweig P-F Study and Test Duration, Socioeconomic Status, and Religion - - - - - - 28
Patterson, C. H. Diagnostic Accuracy or Diagnostic Stereotypy? - - - - - - 483
Pearl, D. See *Armitage, S. G.*
Phillips, Jeanne S. See *Matarazzo, J. D.*
Powell, Elizabeth K. See *Smith, W. I.*
Powers, W. T., and Hamlin, R. M. Relationship Between Diagnostic Category and Deviant Verbalizations on the Rorschach - - - - - - 120
Price, A. C., and Deabler, H. L. Diagnosis of Organicity by Means of Spiral Aftereffect - - 299
Price, J. R., and Thorne, G. D. A Statistical Comparison of the WISC and Wechsler-Bellevue, Form I - - - - - - 479
Pruyser, P. W., and Folsom, Angela T. The Rorschach Experience Balance in Epileptics - 112
Rabin, A. I. See *Hirschstein, R.*
Rafi, A. A. The Discriminative Power of the Strauss-Lehtinen Battery in Adult Mental Patients - - - - - - 135

Raush, H. L. See *Harway, N. I.*
Reid, J. W. See *Bieri, J.*
Reitan, R. M. The Relation of the Trail Making Test to Organic Brain Damage - - - - 385
Remy, R. M. See *Jourard, S. M.*
Rigler, D. See *Harway, N. I.*
Rosenblum, S., Keller, J. E., and Papania, N. Davis-Eells ("Culture-Fair") Test Performance of Lower-Class Retarded Children - - - - 51
Rosenthal, D. Changes in Some Moral Values Following Psychotherapy - - - - 431
Ross, S. See *Smith, W. I.*
Sarbin, T. R., and Hardyck, C. D. Conformance in Role Perception as a Personality Variable - - - - 109
Saslow, G., Matarazzo, J. D., and Guze, S. B. The Stability of Interaction Chronograph Patterns in Psychiatric Interviews - - - - 417
Scherer, I. W., Winne, J. F., and Baker, R. W. Psychological Changes over a Three-Year Period Following Bilateral Prefrontal Lobotomy - - - - 291
Schulz, R. E., and Calvin, A. D. A Failure to Replicate the Finding of a Negative Correlation Between Manifest Anxiety and ACE Scores - - - - 223
Scodel, A. See *Miller, J. S.; Mussen, P. H.*
Sells, S. B. See *Barry, J. R.*
Sersen, E. See *Mayzner, M. S., Jr.*
Severson, W. E. See *Corsini, R. J.*
Shontz, F. C. MMPI Responses of Patients with Multiple Sclerosis - - - - 74
Singer, J. L. Delayed Gratification and Ego Development: Implications for Clinical and Experimental Research - - - - 259
Singer, J. L., and Sugarman, D. A. A Note on Some Projected Familial Attitudes Associated with Rorschach Movement Responses - - - - 117
Sisson, B. D., and Taulbee, E. S. Organizational Activity on the Rorschach Test - - - 29
Smith, W. I., Powell, Elizabeth K., and Ross, S. Food Aversions: Some Additional Personality Correlates - - - - 145
Sonder, Sylvia L. Perceptual Tests and Acute and Chronic Status as Predictors of Improvement in Psychotic Patients - - - - 387
Sontag, L. W. See *Baker, C. T.*
Spielberger, C. D. See *Goodstein, L. D.*
Starer, E. Cultural Symbolism: A Validity Study - - - - 453
Stauffacher, J. C., Marks, J. B., and Ax, A. F. The Effect of Pinhole Observation on the Flicker Fusion Frequency in Patients with Brain Pathology - - - - 150
Stern, J. A. See *Johnson, Laverne C.*
Stone, D. R. Responses to Imagined Auditory Stimuli as Compared to Recorded Sounds - 254
Stotsky, B. A. Predicting Success on the Member-Employee Rehabilitation Program - - 274
Stotsky, B. A. Differential Responses of Normals, Psychoneurotics, and Psychotics on Rorschach Determinant Shift - - - - 335
Stotsky, B. A., and Lawrence, J. F. Various Rorschach Indices as Discriminators of Marked and Little Conceptual Disorganization Among Schizophrenics - - - - 189
Stotsky, B. A. See *Gibby, R. G.*
Strupp, H. H. An Objective Comparison of Rogerian and Psychoanalytic Techniques - - 1
Strupp, H. H. Psychotherapeutic Technique, Professional Affiliation, and Experience Level - - - - 97
Strupp, H. H. The Effect of the Psychotherapist's Personal Analysis upon His Techniques 197
Sugarman, D. A. See *Singer, J. L.*
Sultan, Florence. See *Eron, L. D.*
Summerskill, J., and Darling, C. D. Emotional Adjustment and Dieting Performance - - 151
Taulbee, E. S. See *Sisson, B. D.*

Taylor, D. M. Changes in the Self Concept Without Psychotherapy - - - - - - - - 205
Thomas, R. W. See *Gibby, W. G.*
Thorne, G. D. See *Price, J. R.*
Tolman, Ruth S. See *Meyer, M. M.*
Tresselt, M. E. See *Mayzner, M. S., Jr.*
Trites, D. K. See *Barry, J. R.*
Tunney, T. E. See *Corsini, R. J.*
Turner, C. J. See *Mill, C. R.*
Uehling, H. F. See *Corsini, R. J.*
Vernier, Claire M., Whiting, J. F., and Meltzer, M. L. Differential Prediction of a Specific Behavior from Three Projective Techniques - - - - - - - - - - - - - - - - 175
Wagner, C. M. See *Harris, D. H.*
Waxenberg, S. E. Psychosomatic Patients and Other Physically Ill Persons: A Comparative Study - 163
Webb, W. B. Self-Evaluations, Group Evaluations, and Objective Measures - - - - - - 210
Whiting, J. F. See *Vernier, Claire M.*
Wiener, M. See *Buss, A. H.*
Williams, J. E. See *Goodstein, L. D.*
Windle, C. The Relationships Among Five MMPI "Anxiety" Indices - - - - - - - - 61
Winne, J. F. See *Scherer, I. W.*
Wirt, R. D. Further Validation of the Ego-Strength Scale - - - - - - - - - - - - 444
Wittich, J. J. The Generality of the Prediction of Self Reports - - - - - - - - - - 445
Zelen, S. L. Goal-Setting Rigidity in an Ambiguous Situation - - - - - - - - - - - 395
New Books and Tests - - - - - - - - - - - - - - - - - 75, 154, 235, 318, 404, 488
Index of Authors of Books and Tests Reviewed - - - - - - - - - - - - - - - - - 491

JOURNAL OF CONSULTING PSYCHOLOGY

February, 1955 Vol. 19, No. 1

Contents

An Objective Comparison of Rogerian and Psychoanalytic Techniques: *Hans H. Strupp* - 1

Direction of Perceived Movement in Males and Females: *Walter James Gleason* - - - - 8

Ambiguity as a Therapeutic Variable: *Edward S. Bordin* - - - - - - - - - - - - 9

Differences Between Verbal and Performance IQ's with Mentally Defective Children on the Wechsler Intelligence Scale for Children: *J. Robert Newman and Frank M. Loos* - 16

A Systematic Approach to the Construction of Hypotheses About the Process of Psychotherapy: *Jum C. Nunnally* - 17

Alienation, Social Apperception, and Ego Structure: *Anthony Davids* - - - - - - - - 21

Relationships Between the Rosenzweig P-F Study and Test Duration, Socioeconomic Status, and Religion: *Edward T. Parsons* - - - - - - - - - - - - - - - - - 28

Organizational Activity on the Rorschach Test: *Boyd D. Sisson and Earl S. Taulbee* - - 29

Psychiatric Screening of Flying Personnel with the Cornell Word Form: *John R. Barry, Saul B. Sells, and David K. Trites* - 32

Threshold for the Perception of Human Movement in Inkblots: *Frank Barron* - - - - 33

Factors of Change in Lobotomized Chronic Schizophrenic Patients: *Maurice Lorr, James Q. Holsopple, Richard L. Jenkins, and James P. O'Connor* - - - - - - - - - - - 39

Rorschach Determinant Shift Among Hallucinatory and Delusional Patients: *Robert G. Gibby, Bernard A. Stotsky, Robert L. Harrington, and Richard W. Thomas* - - - - 44

Changes in Wechsler Subtest Scores with Age: *Robert J. Howell* - - - - - - - - - - 47

Davis-Eells ("Culture-Fair") Test Performance of Lower-Class Retarded Children: *Sidney Rosenblum, James E. Keller, and Ned Papania* - - - - - - - - - - - - - - - - 51

A Sexual Deviation Scale for the MMPI: *James T. Marsh, Jessamine Hilliard, and Robert Liechti* - 55

The Relationship of the Wechsler-Bellevue Verbal Scale and the Progressive Matrices Test: *Mahesh M. Desai* - 60

The Relationships Among Five MMPI "Anxiety" Indices: *Charles Windle* - - - - - - 61

Personality Correlates of Certain Religious Attitudes: *William E. Broen, Jr.* - - - - - 64

The Use of Demographic Characteristics in Predicting Response to Hospitalization for Tuberculosis: *Louis J. Moran, George W. Fairweather, Robert B. Morton, and Laurence S. McGaughran* - 65

Results of Treatment of Enuresis by a Conditioned Response Method: *Barclay Martin and Delores Kubly* - 71

MMPI Responses of Patients with Multiple Sclerosis: *Franklin C. Shontz* - - - - - - 74

New Books and Tests - 75

Bureau Edn. [illegible] Research
DAVID HA [illegible] ING COLLEGE
Dated ....
Acs No ....

Brief Reports:

A New Publication Procedure

For a trial period, the *Journal of Consulting Psychology* will accept *"brief reports"* for early publication without expense to the author, under certain conditions. The new procedure is intended to reduce publication lag and permit the circulation of worth-while studies of specialized interest or minor importance which cannot be accepted at present because of limitations of space.

A *brief report,* strictly limited to one printed page, may be accepted if the author agrees to make a more extended report available. Up to four pages in each issue may be devoted to brief reports, published in the order of their receipt without regard to the dates of receipt of articles accepted for regular publication. It is anticipated that most of the brief reports can appear in the first or second issue to go to press following their acceptance.

The procedure for submitting a *brief report* is:

1. The author sends the editor *both* the brief report and a more complete report of the study. The full report should be of sufficient length to give a clear account of the background, procedure, results, and conclusions of the study.

2. The *full report* is prepared in the style specified by the *Publication Manual* (1), except that it may be typed with single spacing. The editor will send the full report to the ADI.

3. The *brief report* gives a clear, condensed summary of the procedure of the study, and as full an account of the results as space permits. To insure that the brief report does not exceed one printed page, the author prepares it according to these specifications:

The text of the brief report, including all matter except the title and the author's lines, does not exceed 80 typewritten lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, or 4.2 inches long in pica. The manuscript is double spaced throughout and, except for the short lines, follows the standard style (1).

Headings, tables, and references are avoided in the brief report. If essential, they must be counted in the 80 lines. The brief report includes, within its 80 lines, a footnote in this style: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. [giving the author's full name and address] or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. ..... from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $.......... for microfilm or $.......... for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

4. The author of a brief report prepares at least 100 mimeographed copies of the extended report. He agrees: (*a*) to send copies of the extended report upon request as long as the supply lasts, and (*b*) not to submit the extended report for publication to another printed journal of general circulation.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.,* 1952, **49**, 389–449.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

An Objective Comparison of Rogerian and Psychoanalytic Techniques[1]

Hans H. Strupp[2]
Washington, D. C.

The research effort by Carl Rogers and his students to describe and elucidate the process of nondirective counseling has provided objective evidence in an area in which such evidence has been sorely lacking. Unfortunately, this impetus has not extended to other theories of psychotherapy, particularly those based more explicitly upon psychoanalytic principles, which today claim a numerically much larger following. The result of this development is a reasonably well-documented body of knowledge concerning client-centered counseling but almost a total absence of empirical data on psychoanalytically oriented therapy (cf. Saslow's recent review [7]).

It is not surprising that much of the published research has focused upon the results of treatment and upon the development of methodologies for the description of the interpersonal process of therapy. It was also inevitable that the patient, for whose benefit the treatment is conducted, would receive a fair share of attention from investigators; it is less apparent why so little research effort has been expended upon the psychotherapist. Yet, it is almost axiomatic that the therapist's personality and attitudes are the prime determiners of the character of his therapeutic operations. On the nondirective side, several investigators have given attention to the therapist's verbal behavior (e.g., Porter 4, Seeman 8, Snyder 9, among others), but with the exception of a few isolated studies, no systematic attempt has been made to study the techniques of psychotherapists not subscribing to the Rogerian frame of reference. The work reported by Fiedler (2) in Mowrer's recent compendium represents an important beginning, but it must be broadened in scope and supplemented by a good deal of down-to-earth exploratory research to give us a better idea of what is meant by such dimensions as "expertness," "ideal therapeutic relationship," etc. In short, as research into the process of psychotherapy increases and as psychotherapy acquires the status of a scientific discipline, we may expect to see a concomitant rise in the number of investigations bearing upon the contributions of the psychotherapist to the treatment situation.

In keeping with this orientation, the present inquiry attempts an empirical comparison between the psychotherapeutic techniques used by Rogerian and psychoanalytically oriented psychologists. The rationale guiding this work may be stated thus: Techniques of psychotherapy are rational and systematic procedures designed to produce an amelioration of the patient's mental health. Every communication by the therapist is ideally placed in the service of the over-all objectives of therapy, as conceived and elaborated in the prominent theories. Techniques may therefore be considered as theory translated into action. Without getting embroiled in discussions about the extent to which Rogers' client-centered theory is basically different from psychoanalytic principles, or whether Rogers' theory is in fact nondirective, it is evident that differences in emphasis do exist among the major theories of psychotherapy. So much is certain. But what about theory in action? Do the operations of psychotherapists proceed in accordance with the theoretical precepts to which they claim allegiance? If so, we would know at least that different theories lead to specific therapeutic procedures predictable from the theory, although the therapeutic effectiveness of any procedure is thereby no further elucidated. If not, theory would have little practical import for therapy.

---

[1] This article is one in a series of investigations concerned with the objective study of psychotherapists' verbal operations. It represents one phase of the author's doctoral dissertation submitted to the Graduate Council of the George Washington University (10), and was in substance presented at the 1954 meeting of the Eastern Psychological Association in New York. Grateful acknowledgment is made to Professor Curtis E. Tuthill, Dr. E. Lakin Phillips, and Leon Salzman, M.D., for valuable suggestions and criticisms.

[2] While carrying out this work, the author held positions successively with the Department of the Air Force and the Department of the Army. Neither organization has any connection with this research.

Investigations of the techniques of Rogerian counselors have shown (e.g., 4, 8, 9) that they make considerable use of the techniques advocated by Rogers, such as reflections of feeling, nondirective leads, etc. With respect to psychoanalytically oriented therapy we have few objective data. It is well known, of course, that the ultimate emphasis of psychoanalytic technique rests upon interpretation, especially of the transference situation in the broadest sense. However, interpretations are not be to made frequently, indiscriminately, or at all in the initial phases of therapy (3, ch. VIII). What techniques, then, does the psychoanalytically oriented therapist use at the beginning of therapy when the patient has as yet produced little information and the therapist is still unclear about the underlying dynamics? By singling out the Rogerian and psychoanalytic systems for the present comparison, we hope to provide objective evidence on the operational similarities and differences of two important theoretical orientations.

Specifically, followers of Rogers' system, in accordance with the principles enunciated in *Counseling and Psychotherapy* (5) and *Client-centered Therapy* (6), should be distinguishable from non-Rogerian therapists in regard to technique in (*a*) a larger number of "reflective" responses, (*b*) fewer interpretations, (*c*) fewer exploratory questions, (*d*) less reassurance, (*e*) fewer instances of passive rejection and outright antagonism. Analytically oriented therapists, on the other hand, should differ from Rogerians in (*a*) offering a larger number of interpretations, (*b*) more explorations, and (*c*) fewer reflective responses.

Finally, it was hypothesized that the variable of personal analysis might disclose systematic differences in verbal techniques.[3]

Method

To study the therapist's verbal behavior under reasonably standardized conditions it was considered advisable to avoid the complexities of the actual therapeutic situation, but rather to have the therapist respond to a carefully preselected sample of patient communications. Thus, a series of 27 short paragraphs of patient statements was culled from published therapeutic interviews and typed on individual cards. A bare minimum of background information preceded each statement. It was stipulated that the materials pertained to early interviews to forestall the possible criticism that the interpersonal relationship between therapist and patient would make it difficult to formulate a response. The 27 statements were regarded to be a fairly representative cross section of verbalizations typically heard from neurotic patients in early interviews. Represented in the series were a variety of complaints, statements by a seriously disturbed near-psychotic patient, suicide threats, and other transference reactions, such as blocking, negativism, requests for direct advice, and open hostility.

Such "miniature therapeutic situations" obviously introduce a certain degree of artificiality and fall short of replicating the actual therapeutic session. While the degree of correspondence between the therapist's actual operations (the criterion) and his performance in a "test" situation should eventually be explored, we may, for the present, grant a certain conditional validity to our experimental procedure.

---

[3] It should be noted that our concern here is exclusively with but a limited aspect of the therapeutic process—the therapist's verbal contribution which is at no time considered coextensive with his total contribution to the treatment situation. But since techniques are described, advocated, taught, and manipulated, it is a legitimate undertaking to inquire into their application and to explore the extent to which they are a function of the practitioner's theoretical orientation.

Table 1

The Sample by Experience Level and Personal Analysis

	Personal analysis			No personal analysis			
Group	Experienced	Inexperienced	Subtotal	Experienced	Inexperienced	Subtotal	Total
Rogerian psychologists	3	2	5	1	2	3	8
Psychoanalytically oriented psychologists	3	3	6	0	1	1	7
Total	6	5	11	1	3	4	15

Procedure and Subjects

Therapists who had indicated a willingness to participate were visited and interviewed individually. Cooperation was surprisingly good, and very few of the therapists who were approached refused. Each interview lasted about one and a half to two hours.

The respondent was presented with the experimental series of patient statements and requested to state what response, if any, he would make to the hypothetical patient.

Information concerning the therapist's training, experience, theoretical orientation, etc. was collected at the end of the session.

The subjects were 15 psychologists engaged in the practice of psychotherapy; 8 were Rogerians and 7 were psychoanalytically oriented. Practically all respondents were attached to clinics in the Washington, D. C. area, and in addition to psychotherapy performed diagnostic and related functions. The data were supplemented by responses similarly collected from 25 psychiatrists and 9 psychiatric social workers, who constituted two psychoanalytically oriented reference groups.

Seven psychologists had more than five years of experience in therapy, and will be referred to as "experienced therapists." Table 1, which presents the respondents' experience level, also indicates that five Rogerians and six non-Rogerian psychologists had undergone personal analysis as part of their training. Rogerians and non-Rogerians appear to be well equated as regards experience level and personal analysis.[4]

Results

Therapists' responses were categorized by Bales' system of interaction process analysis (1), a well-known general purpose framework for describing social interactions. This system commended itself for present purposes because of its theoretical neutrality as regards the conflicting schools of psychotherapy and because of the concentrated effort which has gone into the formulations.

The psychologists' responses yielded a total of 553 score units (mostly simple sentences or thought segments). The reliability of the categorizations was tested by several methods and found to be 78 per cent.[5]

Table 2 shows Bales' 12 categories, together with the responses of Rogerians and non-Rogerians assigned to each. Figure 1 presents the same results in profile form. The *t* values (computed by the formula for uncorrelated percentages) indicate highly statistically significant differences between Rogerian and analytically oriented therapists in almost all applicable categories. A chi-square test for the distributions is significant beyond the .001 level of confidence.

---

[4] A certain degree [illegible] heterogeneity must be assumed among the respondents. There is no "pure" Rogerian just as there is no "pure" psychoanalytically oriented therapist. Most Rogerians must have come in contact with psychoanalytic concepts, particularly those who had undergone personal psychoanalysis. The converse is not necessarily true. Also, while there is considerable diversity in psychoanalytic thinking, most therapists surveyed here (including psychiatrists and social workers) had come under the influence of the Washington School of Psychiatry, and thus espoused neo-Freudian principles and concepts. The classificatory problem was resolved by including all psychoanalytically oriented therapists in one group and all therapists who acknowledged Rogerian influences in the other.

[5] Since data were collected from psychiatrists and social workers for related projects, the reliability was tested upon a stratified random sample of 20 per cent of the cases. The reliability percentage figure represents the average agreement among three judges who independently scored 370 response units.

Figures 2 and 3 attempt a comparison between experienced and inexperienced therapists in each of the two groups. Rogerian therapists show a highly significant decline in reflective responses (Category 6) with increasing experience, and a corresponding rise in exploratory responses (Category 8). The difference in Category 12 appears to be of little practical importance. Chi square is again highly statistically significant. Differences be-

Table 2

Scoring Categories and Response Frequencies

Category no.	Interaction process analysis definition	"Psychotherapeutic" definition†	Rogerians' responses		Non-Rogerians' responses		
			N	%	*N*	%	*t*
1	*Shows solidarity,* raises other's status, gives help, reward.	Gives reassurance, encouragement, shows compassion, tenderness.	8	2.6	13	5.3	1.69
2	*Shows tension release,* jokes, laughs, shows satisfaction.	Not applicable‡	0	0	0	0	
3	*Agrees,* shows passive acceptance, understands, concurs, complies.	Shows passive acceptance, understanding, is permissive (includes silences).	16	5.2	39	15.8	4.10**
4	*Gives suggestion,* direction, implying autonomy for other.	Proposes course of action, defines ("structures") the therapeutic situation.	3	1.0	15	6.1	3.46**
5	*Gives opinion,* evaluation, analysis, expresses feeling, wish.	Interprets, analyzes behavior patterns; thought in process, inferential reasoning.	11	3.6	25	10.0	3.10**
6	*Gives orientation,* information, repeats, clarifies, confirms.	Restates, clarifies, "reflects" (Rogerian response).	231	75.5	35	14.2	14.26**
7	*Asks for orientation,* information, repetition, confirmation.	Asks factual questions, expresses lack of knowledge, uncertainty.	0	0	17	6.9	4.60**
8	*Asks for opinion,* evaluation, analysis, expression of feeling.	Explores, asks for elaboration or expression of feeling (includes open-ended questions, "nondirective leads").	30	9.8	88	35.6	7.37**
9	*Asks for suggestion,* direction, possible ways of action.	Not applicable‡	0	0	0	0	
10	*Disagrees,* shows passive rejection, formality, withholds help.	Shows passive rejection, disbelief; ignores requests or complaints; thwarts, frustrates.	3	1.0	13	5.3	3.14**
11	*Shows tension,* asks for help, withdraws out of field.	Not applicable‡	0	0	0	0	
12	*Shows antagonism,* deflates other's status, defends or asserts self.	Shows antagonism, aggression, sarcasm, irony, cynicism.	4	1.3	2	.8	.56
	Total		306	100.0	247	100.0	

† This column gives examples of the kinds of responses by therapists included in the categories, which were in no way redefined.

‡ The absence of responses in these categories is probably a function of the experimental conditions under which responses were elicited.

** Significant at the .01 level of confidence.

tween experienced and inexperienced analytically oriented practitioners seem to be due to chance.

The final tabulation (Figure 4) shows considerable overlap between Rogerian therapists who had been analyzed and those who had not. Noteworthy differences are observed in Categories 3 and 8: analyzed therapists give significantly fewer responses classifiable as passive acceptance and explorations. Because of the small number of clinicians involved in this comparison, the results must be considered highly tentative. (Chi square was significant at about the .14 level.) A similar comparison for non-Rogerian therapists proved unfeasible for the same reason.

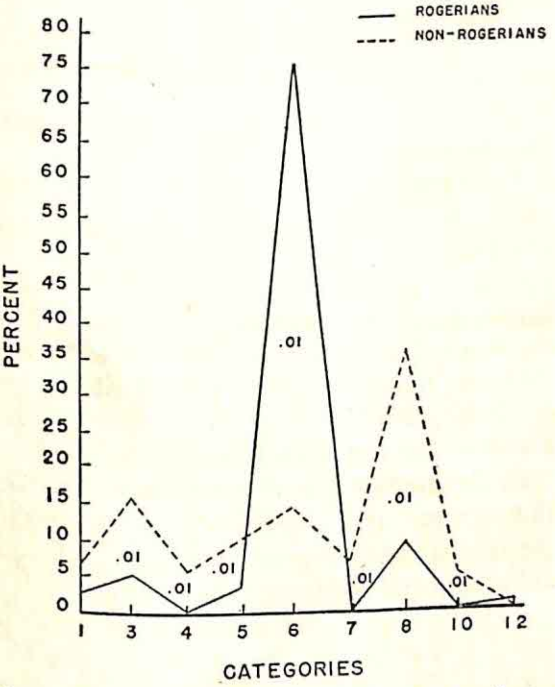

Fig. 1. Response distributions of Rogerian and psychoanalytically oriented psychotherapists.

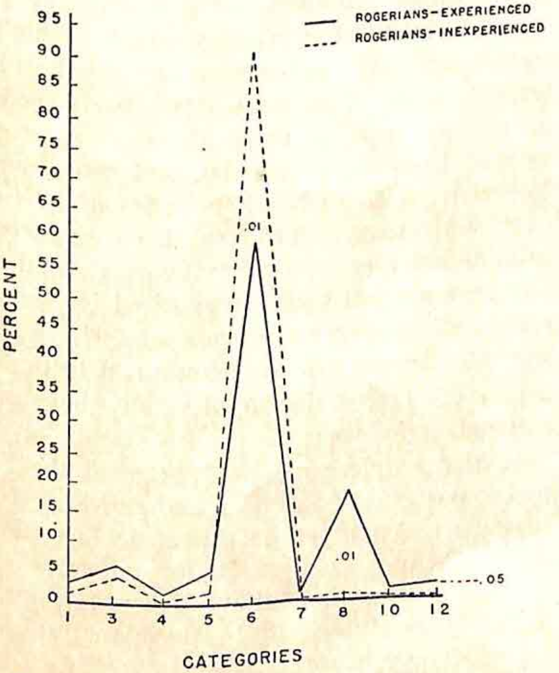

Fig. 2. A comparison of the response distributions of experienced and inexperienced Rogerian psychotherapists.

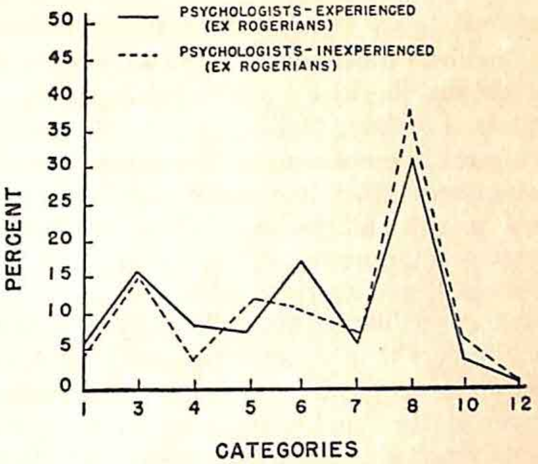

Fig. 3. A comparison of the response distributions of experienced and inexperienced non-Rogerian psychotherapists.

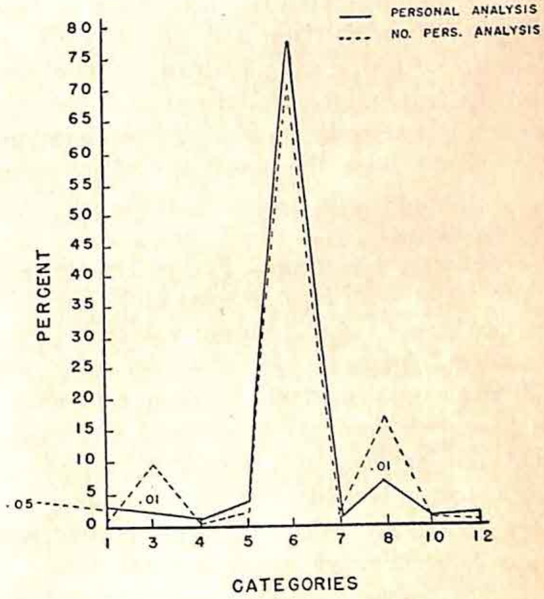

Fig. 4. Response distributions of analyzed and unanalyzed Rogerian psychotherapists.

Discussion

The sharp differences between Rogerian and psychoanalytically oriented psychologists constitute a striking finding, which in general confirms the predictions made from theoretical expositions. Thus, it is noted that Rogerians rely heavily on reflective techniques, with a corresponding neglect for other types of responses.[6] Non-Rogerians, while showing a pronounced predilection for exploratory responses, reveal more than minimal frequencies in such categories as passive acceptance, structuring, interpretation, and possibly reassurance, direct factual questions, and passive rejection. Their responses are distributed more evenly over the range of techniques than those of the Rogerians and, as was shown in another context, closely resemble the response patterns of psychiatrists and social workers. These results, of course, indicate nothing concerning the relative effectiveness of these techniques, but regardless of other considerations there is at least the possibility that undue preoccupation with any one technique may lead to stereotypy.

The data presented in Figure 2 indicate a noteworthy decline of responses in the Rogerians' most favored category, which is not paralleled by non-Rogerians. An analogous difference was observed between experienced and inexperienced psychiatrists. This suggests that an increase in professional experience leads to a diversification of technique, and that, conversely, reliance on one specific technique is a characteristic of inexperience.

The differences between analyzed and unanalyzed Rogerian therapists, exhibited in Figure 4, while based on a small number of cases, are in general agreement with similar analyses for psychiatrists, non-Rogerian psychologists, and social workers. This observation applies particularly to the larger number of responses categorized as passive acceptance (no distinction was made between silent responses and other signs of passive acceptance, such as "Yes," "Mmmh," etc.). The statistically significant difference in exploratory responses was not observed in any other analyses. While only inadequately explained at this time, this finding is of significance here because it clearly cuts across the Rogerian–non-Rogerian cleavage which these analyses have demonstrated.

The results of this investigation leave little doubt that adherence to the Rogerian frame of reference is an important determinant of a psychotherapist's techniques, transcending professional affiliation and experience level. Also, within the limits imposed by the experimental situation, psychoanalytically oriented psychologists, psychiatrists, and social workers form a rather homogeneous group with respect to their therapeutic operations. Future investigations must refine these analyses and specify the conditions under which a given technique achieves a particular therapeutic effect.

Summary

As one in a series of investigations to elucidate the psychotherapist's contribution to the treatment situation, the present study attempts a systematic comparison between the techniques used by Rogerian and psychoanalytically oriented psychologists in early therapeutic interviews. Therapeutic responses were secured from eight Rogerian and seven non-Rogerian psychologists to a series of 27 patient statements which had been extracted from actual therapeutic interviews. A total of 553 response units was categorized by Bales' system of interaction process analysis. Average rater agreement was determined to be 78 per cent. The major results may be summarized as follows:

1. Sharp differences exist between the response patterns of Rogerian and non-Rogerian therapists. As might be expected, Rogerians show a strong predilection for reflective responses, with a concomitant lack of responses in all other categories. Psychoanalytically oriented psychologists distribute their responses more evenly over a variety of techniques although they prefer explorations at

---

[6] It is interesting to note that the present frequency of reflective responses (75 per cent) falls about midway between two comparable figures reported for actual therapeutic interviews. Using a somewhat different classification scheme, Snyder (9) categorized 62.6 per cent of counselor responses as "nondirective," whereas Seeman (8) gave a figure of 85 per cent.

this stage of therapy. The latter group agrees closely with psychiatrists and social workers whose techniques were investigated in related projects.

2. Rogerians disclose a significant decline in reflective responses as the level of their professional experience increases. While non-Rogerian psychologists fail to show a comparable trend, there is other evidence that an increase in experience leads to a diversification of therapeutic technique. Exclusive reliance on one technique appears to be a correlate of inexperience.

3. Analyzed Rogerians, like other therapists whose training has included personal analysis, show significantly fewer silent responses (passive acceptance).

This study thus provides preliminary objective evidence on the techniques of psychotherapists (psychologists) who are following respectively Rogerian and psychoanalytic principles, and documents pronounced differences in therapeutic technique specifically attributable to the clinician's theoretical viewpoint.

Received May 10, 1954.

References

1. Bales, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley Press, 1950.
2. Fiedler, F. E. Quantitative studies on the role of therapists' feelings toward their patients. In O. H. Mowrer *et al., Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 296–315.
3. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univer. of Chicago Press, 1950.
4. Porter, E. H., Jr. The development and evaluation of a measure of counseling interview procedures. *Educ. psychol. Measmt,* 1942, **3**, 105–125, 214–238.
5. Rogers, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
6. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
7. Saslow, G. Psychotherapy. *Annu. Rev. Psychol.,* 1954, **5**, 311–336.
8. Seeman, J. A. A study of the process of nondirective therapy. *J. consult. Psychol.,* 1949, 13, 157–168.
9. Snyder, W. U. An investigation of the nature of non-directive psychotherapy. *J. gen. Psychol.,* 1945, 33, 193–223.
10. Strupp, H. H. An empirical study of certain psychotherapeutic operations: an exploration of the verbal response techniques of psychiatrists, psychologists, and psychiatric social workers. Unpublished doctor's dissertation, George Washington Univer., 1954.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Direction of Perceived Movement in Males and Females[1]

Walter James Gleason
Mental Hygiene Consultation Service, Fort Knox, Kentucky

Rorschach notes a quality of movement on the inkblots which he calls flexion or extension. He states that subjects who see extension movements are active individuals with strong drive toward self-assertion. Those who see flexion movements are thought to be passive, resigned, neurasthenic individuals. Rorschach also quotes Oberholzer as writing that flexion movements in a man represent the feminine part of his sexuality.

Beck has elaborated on this notion and also defines flexion and extension movements more precisely. He defines the extension *M* as centrifugal, as a pull away from the center. Beck calls the flexion *M* centripetal, a pull toward the center. He seems to agree essentially with the Oberholzer concept of extensor and flexor *M* representing respectively the masculine and feminine attitude in the basic personality structure. Thus Beck states that the stance in the fantasy activity may provide evidence supporting the homosexual lead.

To check on the extensions made by Beck and Oberholzer, a series of 28 plates was constructed with the intent to produce figures in which either centrifugal or centripetal movement could be perceived with equal ease. These plates administered to male and female populations could then be used to test whether males would show a predominance of centrifugal movements, and the females, a predominance of centripetal movements. Centrifugal movement was defined as movement away from the geographical center of the plate field, while centripetal movement was defined as either movement toward the center of the plate field or that remained within the center of the field. Many of the designs were suggested by actual Rorschach movement stances, while the remainder were more abstract figures such as contracting-expanding splashes, cubes, circles, and the like.

The plates were made into slides for group administration and an answer sheet was constructed so that upon viewing each plate, *S* was forced into a selection of one of two responses, one representing a centrifugal direction of movement, and the other representing a centripetal direction of movement. The *S*s included 101 males and 87 females from introductory psychology classes at the University of Maryland.

Chi-square analysis revealed that in no case did the male group exhibit a significantly greater number of centrifugal movement responses than the female group. On only two plates were results approaching significance noted. Neither favored the Oberholzer-Beck hypothesis. Thus the fact of biological sexuality has no bearing on the perception of direction of movement on these plates. If the Oberholzer-Beck hypothesis is true, it would seem that more elements are involved than mere direction of movement.

Brief Report

Received August 10, 1954.

---

[1] An extended report of this study may be obtained without charge from Lt. Walter Gleason, 140-F Prichard Pl., Fort Knox, Kentucky, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4454 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Ambiguity as a Therapeutic Variable

Edward S. Bordin
University of Michigan

One of the major issues in psychotherapy is the answer to the question of what makes an interpersonal relationship therapeutic in the sense that the relationship leads to significant personality changes. In an earlier article (3) I have presented and discussed the hypothesis that therapeutic relationships vary in more than one significant dimension. Now I shall pursue this point further.

Before we can proceed, it is necessary to distinguish psychotherapy[1] from what might be called therapeutic phenomena. Many interpersonal relationships which occur in other than a formal therapeutic setting result in personality changes. Interactions with friends, teachers, and others can result in considerable modification of the affective and motivational aspects of behavior patterns. However, these changes have not resulted from the planned and controlled use of these interpersonal relationships. Changes resulting under conditions of ordinary interpersonal relationships can be called therapeutic phenomena to distinguish them from psychotherapy where one person, the therapist, makes use of interpersonal relationships in a conscious and controlled way to lead to changes in the other person's personality. Therapeutic phenomena may be thought of as resulting from the coincidence that one person's natural ways of reacting correspond with the kinds of reactions from people which will lead to personality change in another person. The problem in the theory and practice of psychotherapy is to identify those characteristics of interpersonal relationships which can make relationships therapeutic.

A multidimensional conception of therapeutic relationships is not universally accepted. As my earlier paper suggested, Rogerian theory (11) seems to have assumed that there is a single characteristic of interpersonal relationships, namely, nondirectiveness or permissiveness, which can account for the therapeutic value of these relationships for all combinations of patient and therapist personality. As a matter of fact the directive-nondirective dimension of therapeutic relationships has never, to my knowledge, been subjected to the type of analysis which would permit a test of the tenability of the unidimensional assumption. Aside from this factual issue, it would appear that an assumption of unidimensionality of therapeutic relationships would seem to stem from a relatively simplified conception of personality structure and personality change. This is an assumption which many will be unwilling to accept. If one accepts the proposition that personalities, whether of patient or of therapist, must be analyzed along multiple dimensions in order to be understood and that these dimensions will be relevant to personality development and change, then a quest for multidimensional characteristics of therapeutic relationships is indicated.[2]

Several possible dimensions which might be relevant to the therapeutic value of relationships were suggested in the earlier paper. Since that time, more intensive theoretical analysis accompanied by considerable direct observation of therapeutic relationships through the medium of recordings, typescripts, and supervisory interviews has led to drastic modification of these first tentative statements.[3] The purpose of this paper will be to present a theoretical analysis of one hypothesized dimension, namely, ambiguity. Future papers will present others of the dimensions being formulated.[4]

---

[1] For purposes of this discussion the term "psychotherapy" should be read to refer to all types of situations where personality modification is the goal, i.e., counseling, short-term therapy, and intensive therapy. The respective pairs of terms, "therapist" and "counselor," and "patient" and "client," will be used interchangeably within the pairs.

[2] This position does not preclude the possibility that the distinction between directive and nondirective relationships refers to a particular combination of independent subcharacteristics. We still would want to study *all* of the combinations of characteristics and their contribution to the therapeutic character of interpersonal relationships.

The Concept of Ambiguity

Ambiguity refers to the stimulus characteristics of the therapeutic situation, of which the therapist is the most significant part. As two people interact, each defines himself to the other as a stimulus object to a greater or lesser degree. As a therapist interacts with a patient, he defines himself and the situation both directly, i.e., by direct statement, and, most frequently, indirectly, by the total import of his actions. It is possible to conceive of three especially relevant areas in which these definitions can take place: (*a*) the topics it is appropriate to discuss with him; (*b*) the closeness and other characteristics of the relationship expected; (*c*) the therapist's values, in terms of the goals that he assumes he and the patient should work toward, as well as in more general terms. The therapist may define with different degrees of clarity these three aspects of himself and the therapeutic task. He may define one or two of them quite clearly and leave the others vague or he may vary the three in different combinations and degrees.

By way of topic the therapist may communicate to the patient that he is supposed to talk about his dreams, his job, or his sexual relationships. The area of choice of topic may be even more restrictively defined, e.g., the patient's relationships with his boss, the courses he will be taking next year, or it may be left more ambiguous, e.g., anything that is of personal concern to the patient. The analytic rule "tell me anything that comes to your mind" represents one of the least restrictive and therefore most ambiguous definitions of the appropriate topics for discussion.

The therapist's definition of the closeness of the relationship is undoubtedly influenced by the content of his definitions of appropriate topics. It is customary, in introducing the analytic rule, to talk about the therapeutic situation as different from and not subject to the usual inhibitions associated with ordinary social intercourse. When a therapist indicates that he expects the patient to talk about very personal feelings he implies that it will be a very personal relationship. However, many other aspects of the therapist's behavior are probably more important than formal "structuring" in defining the expected characteristics of the relationship, e.g., the counselor who addresses his client by his first name on first meeting is suggesting an adult-child or fatherly relationship. Whether the therapist permits himself to exchange feelings with the patient or maintains an emotionally controlled objective distance is influential in conveying therapist definitions of the relationship. Similarly, the therapist may be quite communicative of his values or leave them relatively undefined.

The patient reacts to all three aspects of the situation simultaneously. All three aspects are dynamically interrelated parts of an organic whole. Therefore, the effect is to create situations that vary in the degree to which the stimulus field determines the response. In more ambiguous situations different patients will react in different ways. Some will see the therapist as demanding and critical, others will see him as aloof and disinterested, etc. In less ambiguous situations patients will tend to agree more in their perceptions. This last statement must be qualified by considerations of patient personality factors which will be discussed in the next section.

---

[3] I am indebted to the Rackham Fund of the Graduate School, University of Michigan, for making possible the collection of interview protocols which facilitated the process of analysis.

[4] A research grant, M-516, from the National Institute of Mental Health of the National Institutes of Health, U. S. Public Health Service, is making possible an active program of empirical exploration of therapeutic interactions. Drs. A. T. Dittmann (now with NIMH) and H. Raush are coparticipants as principal investigators. Dr. N. I. Harway is research coordinator. Other participants, past and present, include Dr. R. L. Cutler, D. Rigler, and Z. Sperber. All have contributed to the ideas in this paper to some degree; however, the responsibility for this statement is mine. Many of the additional dimensions being formulated are the specific contributions of my colleagues, Drs. Dittmann and Raush.

In effect, we are talking about the degree to which the therapist either advertently or inadvertently gives structure to the stimulus field for the patient. The analyst as a blank screen appears to represent an example of one extreme of the ambiguity-structuredness dimension. Through long periods of silence the analyst minimizes the amount of structure the situation presents to the patient. His admonition "tell me whatever comes to your mind" gives the patient little material for a perception of specific appropriate content for communication. The analyst's position behind the patient eliminates or minimizes the possibility of the patient's obtaining structuring cues from facial expressions or other fragmentary movements. Information-bound counseling relationships, on the other hand, are good examples of extremely structured situations where the purposes of the meeting and the goals are usually both explicitly and implicitly defined in relatively definite terms and where the topic is often restricted by a series of very delimited questions which can be answered "yes" or "no."

To summarize, ambiguity in the context of psychotherapy refers to a quality of the stimulus characteristics of the therapist and the therapeutic task. Ideally, the discrimination of degrees of ambiguity would be based upon the variability of responses among people to a given stimulus complex. Until now, no practical way of applying this direct approach to the measurement of ambiguity in psychotherapy has been developed. However, Osburn (10) did establish two reliable methods of measurement, one based upon ratings of sections of therapeutic interactions ranging in size from a typescript page to an interview unit, the other based on coding of discrete categories of behavior, each category weighted and then all combined to yield a score. These two methods of measurement were found to yield correlations with each other approximately equal to their reliabilities. Further, Osburn presented evidence to support the hypothesis that ambiguity is a unidimensional characteristic of therapeutic interviews. Although Osburn found that ambiguity seemed to operate as a single scale, he also found that his raters were able to make discriminations only in ordinal terms. The reliability of Osburn's coding method of measurement was confirmed as a by-product of Dibner's (5) study of the relationship of anxiety to ambiguity to be discussed later.

Relation to Personality Theory and Measurement

Ambiguity as a variable in personality theory, particularly in connection with emotional factors in perception, is already well known. Ambiguity is achieved by tachistoscopic exposure of stimuli at speeds which provide decreased visual clarity. The variations in perceptions obtained under those conditions are taken as reflections of organized systems of motivations, emotions, and defenses (2, 4).

Similarly, ambiguity as a factor in personality measurement is already utilized in the basic theory of projective techniques.[5] Projective tests have two characteristics. First, they require that the subject react to stimuli which do not have demanding structural determinants of perception. Second, projective tests make possible the identification of the dimension of the stimulus to which the individual has responded and the content of his response. Knowing these it is possible to make inferences about the motivational and emotional structure of the individual or, in usual terms, his personality structure. For a long period enthusiasm for the projective hypothesis has led merely to a multiplication of types of projective situations. More recently there have been potentially more productive studies designed to amplify the projective hypothesis as it applies to personality assessment. One of the natural objects of study is the degree of ambiguity in any test situation. It seems to me that we can learn more by experimental variations within a projective test situation than from an unending process of devising new tests. Weisskopf-Joelson (13, 14, 15) has illustrated this experimental design by her studies of the effects of various degrees

Bureau Edn. Research
TRAINING COLLEGE
Dated ....
Acs No ....

---

[5] For example, see Abt and Bellak (1, pp. 12–18).

of clarity of the human figures on productivity in the TAT. Miller and his students (9) have demonstrated that a projective test must be studied not only in terms of the test stimuli but as a total situation. They have found that such factors as the personality of the examiner and the types of task definitions given the subject, e.g., "this is a test of intelligence," or "this is a test of imagination," influence Rorschach results.

Frenkel-Brunswik (7) has placed particular emphasis on tolerance for ambiguity as an emotional and perceptual personality variable. She has interrelated intolerance of cognitive ambiguity with emotional ambivalence and rigidity. Her analysis is supported by the results of studies carried on by her and her collaborators in the areas of ethnic prejudice, perceptual ambiguity, and rigidity in problem solving.

The Concept of Ambiguity in Therapeutic Theory

Of all therapeutic theories, Freudian theory makes the most complete and the most explicit use of ambiguity. As has been indicated above, the Freudian emphasis on the therapist as a blank screen is a clear description of the therapist presenting himself and the situation as an ambiguous stimulus object. In his discussions of therapeutic technique, Fenichel (6) dealt rather directly with the relationship between ambiguity and the effectiveness of transference interpretation. Although he does not use the term ambiguity, he does point out that transference interpretation will be ineffective and noncommunicative when the therapist has acted in a manner consistent with the patient's transference-determined perceptions of him. He suggests, further, that a therapist cannot be certain that transference reactions are in fact being manifested if he does not eliminate the reality basis for the patient's reaction. Only thus can he demonstrate to himself as well as to the patient that the patient's reactions to him are controlled by irrational, infantile feelings.

When we turn to nondirective theory, we find that being "nondirective" means, in part, not defining or imposing your own values on the client. Also, the nondirective counselor is supposed to avoid "directive" leads. A directive lead seems to be a demand for the client to talk about a very specific topic in restricted terms as contrasted with a nondirective lead which is a more general invitation to communicate, sometimes including a relatively unrestricted designation of a topic. One gets the impression that Rogerian therapists are much less ambiguous than orthodox analytic therapists. This impression is based on the fact that Rogerian therapists seem to talk more, particularly in the early stages of therapeutic relationships. It is relatively rare for the Rogerian therapist to respond with silence to a patient's responses. It seems as though the Rogerian goal of understanding the client or patient becomes a command to attempt to put into words what the patient is trying to communicate whenever the patient comes to a stop. That this is a developmental trend in nondirective therapy is suggested by the results of Seeman's comparison of nondirective therapy ten years earlier with present-day illustrations (12). One of his findings was that a greater proportion of current nondirective therapists' responses could be characterized as clarification and a smaller proportion of them could be characterized as simple acceptance. According to the meaning of these categories it would seem that nondirective therapists have tended to increase the amount of effort expended in verbalizing patients' feelings and have decreased the frequency with which they respond with relatively simple listening responses. This might lead to the possible conclusion that nondirective therapists are less ambiguous now than they were ten years ago, as contrasted with the nondirectivists' interpretation that they are understanding their clients better.

Application to Therapeutic Theory and Research

The preceding discussion has been designed to demonstrate that ambiguity has meaningful relationships to personality theory and measurement and to therapeutic theory. In this section I shall attempt to show that this construct leads to further expansion and systematization as well as to the formulation of testable hypotheses about therapeutic work.

Let us summarize and make more explicit the fundamental proposition which can be derived from therapeutic and personality theory, namely, that ambiguity in therapeutic relationships serves three major functions. First, it capitalizes on the principle that ambiguous stimuli elicit from people those responses which are most heavily laden with the unique aspects of their life history. In other words, people invest ambiguous stimuli with their own motivational and emotional life. This function of ambiguity makes it possible for the patient, no matter how well oriented to reality, to bring into the therapeutic relationship his major conflictual feelings no matter how unaware he is of them. This, we think, is identical with the Freudian concept of transference so that ambiguity is one of the facilitating factors in the appearance of transference phenomena in therapeutic relationships. Second, by leading to this investment of the patient's motivational and emotional structure, ambiguity makes it possible for the therapist to understand more fully and more deeply the mainsprings of the patient's action. Third, by being ambiguous, the therapist provides a background against which the patient's irrational feelings will be more clearly etched and therefore more readily brought to awareness. Ambiguity helps to insure the effectiveness of the well-timed interpretation.

I shall outline four different areas for which the concept of ambiguity can lead to further clarification of therapeutic phenomena and point the way to significant therapeutic research.

1. The attempt to master the complexity of therapy makes us susceptible to reductionistic approaches such as would be reflected in ruling that all relationships to be therapeutic must be ambiguous for the patient. But the functions of ambiguity in therapeutic relationships must be balanced against our conceptions of the role of anxiety. While anxiety is considered a necessary ingredient of effective therapy, our present knowledge leads us to believe that there is an optimal level of anxiety for each patient and that as the level of anxiety begins to exceed that point, the patient starts to be overwhelmed by his anxiety and eventually all of his energies are consumed in self-preservative efforts, leaving no energy for therapeutic movement.

Psychoanalytic personality theory provides a basis for expecting that anxiety will be positively related to ambiguity. This theory assumes that people try to defend themselves against anxiety associated with conflictual impulses by denying to awareness or distorting those stimuli associated with the impulse. Where the stimuli are ambiguous, the discriminative processes necessary for defense are hampered and greater anxiety will result. Dibner (5) found evidence to confirm the deduction that anxiety will be positively related to ambiguity in an interpersonal relationship. He applied Osburn's (10) methods of measuring ambiguity to admission interviews with patients referred to the psychiatric service of a general hospital. His interviewers, clinical psychologists, carried out half of their interviews with a set to be ambiguous and the other half with a set to be structured. Their compliance with the instructions was confirmed by measures of ambiguity applied to their interviews.

From this confirmed hypothesis of a positive relationship between ambiguity and anxiety one can move to the further deduction that the more intense the person's conflict and the weaker his discriminative and defensive capacities, the greater will be the anxiety experienced at any given level of ambiguity. In other words, patients, varying in their personality structure and particularly in the flexibility and effectiveness of their integrative processes, will vary in the amount of threat and anxiety experienced in ambiguous interpersonal relationships. Thus, it becomes necessary for the therapist to relate the degree of ambiguity to the level of anxiety which will be optimal for the particular patient. Perhaps, one of the critical factors distinguishing effective therapeutic work with schizophrenic patients from work with neurotics is the lesser degree of ambiguity which is demanded in relationships with schizophrenics. This distinction would be consistent with the assumption that schizophrenic personalities are characterized by inadequate and weak ego-integrative abilities. The schizophrenic needs help and encouragement to increase his contact with reality. Even when presented with relatively unambiguous stimulus situations, he is likely to react in purely idiosyncratic personal terms. Fromm-Reichmann (8) has taken a similar position when she holds that the analytic free-association process was inappropriate for schizophrenic patients.

In looking at ambiguity as a factor in the therapeutic relationship, we must not overlook the fact that the therapist may also be subject to its anxiety-provoking effects. In many cases, the structuring of the relationship comes from the therapist's anxiety and discomfort in too free a relationship, one in which he is not able to control and foresee the exact direction of the patient's reactions.[6]

2. As was suggested in our preceding discussions of the functions of ambiguity, we can hypothesize that the intensity of the transference will be a function of duration and degree of the ambiguous aspects of the relationship. This hypothesis must be modified by consideration of the effects of the patient's personality structure on his susceptibility to transference. The more highly integrated the person and the greater his consequent ability to react realistically, the more linear will be the relationship between ambiguity and intensity of transference.

3. There has been a great deal of confusion about the degree of identity between the terms "counseling" and "psychotherapy." Counseling has been applied to therapeutic work with relatively restricted goals with normally integrated clients. Restriction in therapeutic goals usually means limiting intensity of transference. Further, normally integrated clients are usually subject to only minimal anxiety. Therefore, they are less ready to enter a situation in which the act of entering and the conditions of the task imply admissions of personal inadequacy. The nature of the extremely ambiguous situation is such that, in effect, the therapist is saying, "do not try to make the usual discriminations; just speak out in response to your impulses." The implication is that the therapist will protect the client from harm. This is almost a childlike relationship which the normally integrated minimally anxious client will be unwilling to accept. His slight incapacities will not provide sufficient motivation to compensate for the loss in self-esteem in accepting such a relationship. It is likely that this characteristic of less ambiguity will take the form of focusing reactions by the counselor designed to help the client restrict self-exploration and self-expression to those areas of emotion and motivation which are specific to a given problem situation, e.g., making a vocational decision. These two factors then, the need to control transference and the kind of relationship acceptable to the counseling client, suggest that counseling relationships will be characterized by less ambiguity than is true in psychotherapy. Counseling and psychotherapy have been difficult to differentiate. If ambiguity is, in fact, one variable on which they can be differentiated, it will still be difficult to differentiate them dichotomously because therapeutic relationships will probably not distribute themselves in a clearly discontinuous manner.

4. On first analysis, I expected that psychotherapy would be characterized by decreasing ambiguity as the relationship progressed. This expectation was based on the notion that even where the therapist imposes a strict free-association process, eventually he does begin to communicate interpretively to the patient, thus giving greater structure to himself and his reactions to the patient. Inevitably with time, the patient has the opportunity to obtain cues about the therapist and his expectations which define him more concretely. However, Osburn (10) obtained negative results when he tested this hypothesis in studying a sample of counseling relationships. One way to interpret these results is that a revision in theory is indicated. However, these negative findings do not in themselves invalidate our first hypothesis. The negative findings could be due to the fact that this hypothesis is not applicable to counseling situations, but only to those situations characterized by greater ambiguity. If much more intensive therapeutic situations were subjected to analysis, we might then expect to find that the interpretive process has served to structure the situation to an increasing extent as therapy progressed.

Even though the above rationale retains the original prediction about the trend of ambiguity in therapy, the further examination of the issue occasioned by Osburn's results has led to certain additional hypotheses. While to an observer the stimulus character of the therapeutic situation becomes increasingly structured as a function of increased interpretive effort, the dynamics of the situation may be such that, for the patient, the ambiguity has not decreased. When a therapist's interpretation is experienced by the patient as an increase in intensity of his awareness of himself, it will be the patient, himself, as a stimulus object which becomes more structured, not the therapist. When the therapist points out to the patient the many ways in which he is defending himself against passive impulses, the patient experiences this as, "Yes! I feel dependent, I want to be dependent and I am afraid of these feelings," rather than, "He thinks I am dependent and that I am afraid of these feelings." Such will be the stimulus situation for the patient when he is at a stage where he is rushing ahead toward self-discovery. But there are also stages where the approaching awareness is freighted with irrational fear and the patient is actively seeking to avoid it. At these times the confronting type of interpretation, even when effective, is likely to be experienced as indicating to the patient what the therapist thinks of him rather than as a characteristic of himself, thus structuring the situation. Similarly, inappropriate interpretations will have the same effect.

---

[6] This hypothesis is to be tested in a study planned by David Rigler.

Summary

In this paper, the concept of ambiguity as one characteristic of interpersonal relationships relevant to their therapeutic effect has been outlined. Ambiguity is a quality of the stimulus characteristics of the therapist and the therapeutic task which permits patients to vary their perceptions and responses to the therapist as a function of their personality structures. Methods of measuring this quality have been developed and applied in research on psychotherapy. The previous explicit and implicit uses of this concept in personality theory and measurement and in theories of psychotherapy have been reviewed. Finally, predictions were made about the functional relationship between this variable and therapy with schizophrenics, the experience of anxiety, the distinctions between counseling and psychotherapy, and about its relationships to interpretation.

Received May 21, 1954.

References

1. Abt, L. E., & Bellak, L. (Eds.) *Projective psychology; clinical approaches to the total personality.* New York: Knopf, 1950.
2. Blake, R. R., & Ramsey, G. V. (Eds.) *Perception—an approach to personality.* New York: Ronald, 1951.
3. Bordin, E. S. Dimensions of the counseling process. *J. clin. Psychol.,* 1948, **4**, 240–244.
4. Bruner, J. S., & Krech, D. (Eds.) *Perception and personality.* Durham: Duke Univer. Press, 1949.
5. Dibner, A. The relationship between ambiguity and anxiety in a clinical interview. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
6. Fenichel, O. *Problems of psychoanalytic technique.* (Trans. by D. Brunswick.) New York: Psychoanalytic Quarterly, 1941.
7. Frenkel-Brunswik, Else. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.,* 1949, **18**, 108–143.
8. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univer. of Chicago Press, 1950.
9. Miller, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.,* 1953, **48**, 367–375.
10. Osburn, H. G. An investigation of the ambiguity dimension of counselor behavior. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
11. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
12. Seeman, J. A study of the process of nondirective therapy. *J. consult. Psychol.,* 1949, **13**, 157–168.
13. Weisskopf, Edith A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.,* 1950, **29**, 407–416.
14. Weisskopf, Edith A., & Dunlevy, G. P. Bodily similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.,* 1952, **47**, 441–445.
15. Weisskopf-Joelson, Edith A., & Lynn, D. B. The effects of variations in ambiguity on projection in the Children's Apperception Test. *J. consult. Psychol.,* 1953, **17**, 67–69.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Differences Between Verbal and Performance IQ's with Mentally Defective Children on the Wechsler Intelligence Scale for Children[1]

J. Robert Newman
University of Illinois

and Frank M. Loos
Lincoln State School

The main purpose of this study was to gather evidence on whether or not there are differences between VIQ and PIQ for mentally defective children.

Three groups of mentally defective children, classified as familial ($N = 128$), undifferentiated ($N = 75$), and brain damaged ($N = 28$), were selected for study. The children selected ranged in age from 7 to 16 years, had Full Scale IQ's of 50 or higher, and showed no evidence of severe motor defect such as paralyses, tremors, etc. All children were in state institutions in the State of Illinois and were administered the WISC by staff psychologists.

Two hypotheses were tested: (*a*) There is no difference between Verbal and Performance IQ's for any of the three groups. (*b*) There are no differences between group psychometric patterns formed by the VIQ and PIQ.

Statistical analyses of the WISC subtest scores yielded results from which the following tentative conclusions can be made:

1. Mentally defective children classified as familial perform significantly higher as a group on the Performance subtest than on the Verbal subtest. The mean IQ difference was 8.07.

2. Mentally defective children classified as undifferentiated also performed significantly higher as a group on Performance than on the Verbal subtest, but the mean difference (4.8) is not as great as for the familials.

3. Mentally defective children classified as having some kind of brain damage due to infection or birth trauma and showing no evidence of severe motor defect showed no difference between VIQ and PIQ.

4. Mentally defective children with brain damage of the type described above perform uniquely on the WISC as a group to the extent that they obtain a higher VIQ and lower PIQ than the undifferentiated group.

The above conclusions refer to group differences and a word of caution is necessary for those who wish to consider the significance of a difference score for an individual. By making use of the mean reliabilities of two tests and the intercorrelations of these tests it is possible to determine what percentage of difference scores is real, i.e., due to nonchance factors. For the WISC Verbal and Performance tests this percentage is 42. The implication should be clear. Less than half of any differences found between these two tests will be true (reliable), nonchance differences. Obviously when considering the individual case only the largest of differences should be considered trustworthy, and small differences should be attributed to errors of measurement.

Brief Report
Received September 17, 1954.

---

[1] An extended report of this study may be obtained without charge from J. Robert Newman, Department of Psychology, University of Illinois, Urbana, Illinois, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4453 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

A Systematic Approach to the Construction of Hypotheses About the Process of Psychotherapy

Jum C. Nunnally[1]
University of Illinois

Most hypotheses about psychotherapy come from the therapist's recollection of the process in individual cases. This procedure has the weak points of being dependent on memory and lacking in statistical aids to analysis. The method presented here offers a means of codifying the therapist's impressions at various points in the therapy. Trends in the data can be easily charted and it is possible to obtain factors and relations among factors of which the therapist himself is not aware. The method deals with the therapist's impressions and in that sense is subjective. The purpose of the method is to generate some operational hypotheses about the process of psychotherapy that can be verified in other experiments.

If we assume that the therapist judges some client behaviors to be more important or indicative than others, and that these "important" behaviors have a generality from case to case, a "universe" of relevant indicators could be defined for the individual therapist. Samples from the universe of behaviors could be treated statistically. Stephenson (2, 3) has developed a method for working with such samples. The method requires that the sample be rated along a continuum relative to an "instructional set." The scores obtained are susceptible to the ordinary statistics of correlation and t test and to the elegancies of factor analysis and analysis of variance.

The method being recommended would start with a concrete therapy case—any case would do. The therapist's collection of "important behaviors in therapy" would be represented in a sample. These could be such as "acted casual and indifferent about his problems," "asked many questions about my personal life," and "spent much time in thanking me for the help." A standard sample could be made for the more general approaches, such as Freudian, client-centered, and Adlerian; but logically each therapist's particular notions could be represented.

The logic of sampling stimuli, in this case statements about therapy, is similar in principle to the more conventional sampling of people. First, it is necessary to define the sampling unit and the universe to be sampled, here Rogers' notions about therapy (1). If there is a formal theory involved in the experiment, it can be represented by "effects" and "level" in factorial design. When there is no formal theory involved in the experiment, as in the present case, the sampling follows principles of randomization. The first step in this procedure is to make an exhaustive collection of the statements in the specified universe. This larger collection of statements is then randomly sampled in order to obtain a manageable number for the experiment. All of the conventional sampling statistics then apply to the relations between the experimental sample and the parent collection of statements. If the parent collection of statements has been sufficiently exhaustive, there is reason to believe that the experimental sample will be representative of the defined universe of statements.

The experimental design would require the therapist to describe the therapy at regular intervals with the trait sample. At the conclusion of therapy, the obtained Q-sort descriptions could be intercorrelated and factor analyzed. The nature of the factors and the pattern of loadings would provide a sounder basis for theorizing about the therapy process than previous procedure.

---

[1] Grateful acknowledgment is given to Dr. Carl Rogers for his part as therapist in the experiment and for his help in developing the ideas presented.

The example to be described was taken from the case of a young woman who participated in 47 interviews of client-centered therapy. The trait sample was collected from *Client-Centered Therapy* (1), largely from Chapter 3. The 60 statements[2] obtained were balanced positively and negatively to represent the appearance of an important behavior and its opposite. The following are examples of the statements included in the sample:

1. Feeling confused about what others really think of her.
2. Wanting to impress and please therapist.
3. Wondering how change can come about.
4. Struggling to express herself.

Soon after the therapy hour, the therapist was asked to describe the client's behavior by distributing the trait sample along a nine-point continuum. Each of the nine points on the continuum was designated by a category number, ranging from 0 to 8. A forced distribution was required for rating the 60 traits such that they conformed to a quasi-normal curve. The frequencies required at each of the nine points on the continuum were 2, 3, 7, 11, 14, 11, 7, 3, and 2, respectively. Stephenson (2, 3) gives a more complete description of this rating procedure.

The two statements placed in pile 8 were deemed most descriptive of the particular therapy interview, the three statements placed in pile 7 were next most descriptive of the interview, and so on down the continuum to the two statements in pile zero that were thought to be the opposite of what took place. Statements placed in the middle of the continuum were neither characteristic nor uncharacteristic of the interview.

The therapist described the therapy on eight occasions,[3] at interviews 1, 4, 7, 11, 15, 25, 38, and 47. The eight descriptions, each a variate in the experimental design, were intercorrelated[4] and factor analyzed (see Tables 1, 2, 3, and 4). The two factors obtained were

Table 1

Intercorrelations Among Sorts

Sort	1	2	3	4	5	6	7	8
1	—							
2	18	—						
3	−23	51	—					
4	−08	61	65	—				
5	−21	35	64	56	—			
6	−45	01	54	28	54	—		
7	−17	39	53	56	32	44	—	
8	−54	−21	36	02	32	72	29	—

Note.—Decimal points omitted.

Table 2

Centroid Factor Matrix*

Sort	I	II	h^2	est. h^2
1	−39	50	40	43
2	43	67	63	64
3	84	18	74	73
4	70	52	76	74
5	69	06	48	50
6	74	−41	72	72
7	63	11	41	44
8	55	−67	75	71

Note.—Decimal points omitted.
* Thurstone's (5) method of centroid factor analysis was used.

---

[2] A more complete study with the method being described would necessitate the use of a larger sample of traits. One hundred statements would more nearly meet the requirements.

[3] The limited number of sorts was necessitated by restrictions on the therapist's time and by choosing interesting sessions to describe. A substantive experiment with this method would require a dozen or more sorts.

[4] Because of the use of a standardized distribution for all of the sorts, the product-moment correlation coefficient can be taken as follows: $r = 1 - (\Sigma d^2/2N\sigma^2)$. The symbol d stands for item differences in the two variates being correlated. Because $2N\sigma^2$ is a constant for all of the sorts made with a particular distribution, a table can be constructed for transforming any Σd^2 directly into a correlation coefficient. With these convenient procedures, correlation coefficients can be obtained at the rate of 20 to 30 an hour.

Reliability estimates for the Q sort can be made in several ways. A duplicate set of statements can be formed and these can be correlated with the experimental sample on the same occasion or after a period of time. The reliability estimate can be made by simply repeating a particular sort with the same sample after a period of time. Often it is easier to estimate the reliability from the matrix of intercorrelations. It is a safe assumption that the reliability will be at least as high as the median communality.

Table 3

Distribution of Second-Factor Residuals*

r	Frequency
00	1
01	7
02	5
03	4
04	4
05	4
06	1
07	1
.	
.	
.	
12	1

Note.—Decimal points omitted.
* Signs being ignored.

Table 4

Rotated Orthogonal Factor Matrix

Interview	Sort	I′	II′
1	1	−28	−55
4	2	53	−57
7	3	85	−03
11	4	77	−37
15	5	67	07
25	6	65	53
38	7	63	00
47	8	42	75

Note.—Decimal points omitted.

placed through one orthogonal rotation[5] to "sharpen" the relationships made apparent by the factors. Each factor can be considered an "effect" with the loadings measuring the extent to which the therapist judged each effect to dominate the separate interviews.

From factor scores[6] obtained for each of the traits, the following traits are found to be high, positively and negatively, in factor I′:

Factor I′

Most Characteristic

Experiencing ambivalent attitudes toward self.
Experiencing the therapist as accurately understanding.
Experiencing warm rapport with therapist.
Feeling relieved to have an hour of her own to just do what she likes.
Wondering who and what she really is.
Experiencing the therapist as interested and wanting to understand.
Feeling skeptical of the "validity" of her own experience.

Most Uncharacteristic

Feeling annoyed with therapist.
Feeling interest and concern for the therapist as an individual.
Feeling that the whole effort is futile.
Worrying about how the therapist sees her.
Feeling that others should change to make her lot a more comfortable one.

The items characteristic and uncharacteristic of factor II′ are as follows:

Factor II′

Most Characteristic

Experiencing real happiness.
Perceiving self from values rooted in primary experience.
Experiencing comfort in being her real self.
Feeling pride in assuming responsibility for self.
Working from new insights to reoriented behavior.
Experiencing progress in therapy.

Most Uncharacteristic

Wondering what the therapist expects of her.
Perceiving self in terms of values introjected from others.
Feeling difficulty in talking about self.
Feeling that she just isn't capable of facing her problems.
Wanting the therapist to assume the burden of effort—wanting key to problems.

Factor I′ could be interpreted as a client-therapist interpersonal activity concerned with the development of trust, faith, and warm emotions. Factor II′ is more intrapersonal for the client, aimed at free exploration and self-trust. It is interesting that the therapist sees the client at her happiest when factor II′ predominates, and ambivalent about herself when factor I′ is the strongest influence. The therapist apparently does not think of the warmth in factor I′ as directed toward himself as an individual. Note the statement "feeling interest and concern for the therapist as an individual" appearing as the second most uncharacteristic trait in factor I′.

---

[5] In connection with the rotational problem here it is interesting to note that principles of "simple structure" do not apply to factor matrices resulting from the use of this method. It would be difficult to conceive of the variates here being segregated in clusters in the vector space. This would mean that the client moves along under one set of conditions and then suddenly jumps to a wholly different set of reactions. The solution found here, where the variates are distributed continuously through the vector space, is more likely.

[6] Factor arrays are obtained by weighting the variates by the formula $\frac{\text{Loading}}{(1\text{-Loading})^2}$ (4, Appendix XX) and averaging the item scores across variates.

Both factors are in the negative at the beginning of therapy: there is neither interpersonal nor intrapersonal "good-will." Factor I′ becomes positive on the second sort (the fourth interview), reaches a maximum on the third sort, and gradually declines to the close of therapy. The implication is that behaviors relative to interpersonal warmth grow and then decline, perhaps necessarily so in the successful therapy.

Factor II′ is less steady in its trend. It is not until the sixth sort that the factor is significantly positive. It becomes zero again in the seventh and rises to its maximum in the last interview. Rising later than factor I′, factor II′ can be construed as dynamically closer to the conceived basic process of therapy. Factor I′ becomes significantly positive in the fourth interview (second sort), but it is not until twenty-one interviews later that factor II′ comes into play.

The analysis raises the following sorts of questions:

1. What behavior on the part of the therapist is necessary to bring each of the factors about?

2. Would it be possible to have factor I′ in the negative and factor II′ positive (would it be possible for the client to actively dislike and distrust the therapist, yet freely to explore himself)?

3. Is happiness a necessary part of factor II′ and ambivalence a necessary part of I′ or can these relationships be better explained by the expectations of the therapist?

4. Could the prolonged and undiminishing preocupation with factor I′ prevent the self-trust and self-exploration deemed necessary for therapy?

5. Would these factors arise in other therapies done by therapists of different orientations? Would the differences be reflected linguistically (as shown by the "important" behaviors included in the particular therapist-sample) more than in the nature of the factors obtained?

Summary

A method has been shown for systematically studying the therapist's impressions about the process of psychotherapy. It entails a *Q*-method (2, 3) treatment of samples of "important behaviors in therapy" taken from current theoretical works. The approach allows the therapist's observations to be conveniently recorded and analyzed by appropriate statistics. The purpose of the method is to generate some more operational hypotheses about the therapy process than presently exist. One case is analyzed to show the method at work.

Received May 17, 1954.

References

1. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
2. Stephenson, W. Some observations on *Q* technique. *Psychol. Bull.*, 1952, **6**, 483–498.
3. Stephenson, W. *The study of behavior.* Chicago: Univer. of Chicago Press, 1953.
4. Spearman, C. *The abilities of man.* New York: Macmillan, 1927.
5. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Alienation, Social Apperception, and Ego Structure[1]

Anthony Davids
Harvard University

Examination of the recent psychological literature bears witness to the fact that an increasing amount of research attention is being devoted to the study of social perception (1, 2, 3, 4, 5, 6, 7, 12, 13, 15, 16). This trend in research interest would seem to be a natural concomitant of the emphasis placed on interpersonal relations in contemporary personality theory and psychiatric practice. Without doubt, most investigators of personality would agree that effectiveness of social relations is an important component of general environmental adjustment, and that accurate social perceptions should be conducive to satisfactory social relations. Despite agreement in regard to the importance of this phenomenon, commonly termed *social perception,* at the present time the psychologist is faced by a multiplicity of unsolved problems concerning relations between personality attributes in the perceiver and his evaluative interpretations of other persons in his social environment. The present investigation represents an attempt to contribute to further understanding of these relations.

More specifically, the twofold purpose of this study is to develop an objective technique for investigating personality and social apperception, and to test a series of predictions pertaining to relations between subjects' ego structure, their apperceptions of the world, and their estimates of other persons' apperceptions. Throughout this presentation, the term *apperception* will be used in place of the more common term *perception*. By use of this concept we do not wish to give the impression that we are exploring some novel area of investigation. The focus of this research is definitely on the problem area that in the bulk of the literature is labeled *social perception*. Our choice of the concept *apperception* is in accord with the distinction between perception and apperception made by Murray (9), in which the latter term is used to designate the process whereby meaning is assigned to a physical stimulus, and includes such things as understanding, interpretation, and classification. In the present study apperception refers specifically to a perception involving an evaluative interpretation of situations and people in the perceiver's world.

Method

Subjects

The subjects were 20 Harvard undergraduates who volunteered to participate in an intensive study of personality. These subjects, selected from a large introductory course in psychology, were paid for their services at the current student employment rate. They were selected with the goal of obtaining a group of young men with varied backgrounds, interests, abilities, and personalities.

The Affect Questionnaire

Description. The Affect Questionnaire was designed to reveal three things in regard to certain critical dispositions of personality: (*a*) the relative amount of the disposition in the subject's personality; (*b*) the relative amount of the disposition he apperceives in

---

[1] This investigation, conducted at the Harvard Psychological Clinic under the direction of Dr. Henry A. Murray, was facilitated by the Laboratory of Social Relations. It was supported in part by a research grant from the Rockefeller Foundation, and grant M-700 from the National Institute of Mental Health, U. S. Public Health Service. The author wishes to express his appreciation to Dr. Murray, who collaborated in the design and construction of the Affect Questionnaire, and provided the clinical evaluations of ego structure.

search
6.3.81

the personalities of other people in his social environment; and (*c*) his evaluation (positive-negative) of the given disposition. The eight dispositions measured by this instrument are: sociocentricity, optimism, trust, egocentricity, pessimism, distrust, anxiety, and resentment. Of these dispositions, the first three are "positive" or socially desirable, and the last five are "negative" or socially undesirable.

There are three forms of the Affect Questionnaire, which are identical in content but differ in instructions. The content consists of ten statements designed to relate to each of the eight dispositions, making a total of 80 items. For each disposition a great many statements indicative of evaluative apperceptions of the world and its people were written; some were original constructions by the experimenters, and others were culled from sources such as Stevenson's *The Home Book of Quotations* (14), and Murray's *Explorations in Personality* (10). From this host of items we selected what we believed to be the ten best items for tapping each disposition from a variety of approaches. The ten items representative of each of the dispositions were then mingled in a random, but systematic, order to constitute the final 80-item questionnaire.

Following are the first eight statements as they appear on the Affect Questionnaire:

1. The most profound happiness is reserved for those who are capable of selfless dedication to a cause.
2. No longer can a young man build his character and his hopes on solid ground: civilization is crumbling, the future is dreadfully uncertain, and his life hangs by a thread.
3. People will be honest with you as long as you are honest with them.
4. A man who never gets angry at anything or anyone is not likely to be treated with respect.
5. There are days when one awakes from sleep without a care in the world, full of zest and eagerness for whatever lies ahead of him.
6. Beneath the polite and smiling surface of man's nature is a bottomless pit of evil.
7. There are times when it is absolutely necessary to use other people as tools in the accomplishment of a purpose.
8. The real substance of life consists of a procession of disillusionments, with but few goals that are worth the effort spent in reaching them.

Administration. The instructions require the subject to rate each statement by marking the appropriate point on a 6-point scale, running from "strongly disagree" to "strongly agree." He responds to these statements under each of the three following conditions considered in this order: (*a*) the way he thinks the "average Harvard student" would rate them; (*b*) the way he himself rates them; and, finally, (*c*) the way he thinks his "ideal person" would rate them. In other words, first the subject is presented with a copy of the questionnaire with mimeographed instructions to respond to each statement the way he thinks the average Harvard student would respond to it. When he has completed this task, he is presented with another copy of the same questionnaire, only this time the instructions state that he is to indicate his own

Table 1

Rank-Order Correlations (Rho) Between Scores on the Eight Dispositions on the Affect Questionnaire ($N = 20$)

Disposition	Optimism	Pessimism	Trust	Distrust	Socio-centricity	Ego-centricity	Resentment
Optimism							
Pessimism	−.47*						
Trust	+.43*	−.64***					
Distrust	−.41*	+.84***	−.82***				
Sociocentricity	+.29	−.40*	+.80***	−.62***			
Egocentricity	−.10	+.76***	−.64***	+.75***	−.52**		
Resentment	−.09	+.59***	−.38*	+.64***	−.57***	+.54**	
Anxiety	−.29	+.82***	−.54**	+.86***	−.42*	+.67***	+.58***

* Significant beyond the .05 level for a one-tailed test.
** Significant beyond the .01 level for a one-tailed test.
*** Significant beyond the .005 level for a one-tailed test.

personal feeling of agreement or disagreement with each statement. Finally, he is presented with a third copy of the same questionnaire, and this time he is instructed to, "Rate the following statements as you believe the *ideal person* would rate them. The ideal person may be real or mythical, living or dead, or purely imaginary. This person should be someone admirable in your eyes, someone who possesses the attributes which you would most like to have in your own personality."

Scoring. Separate scores were computed from responses elicited under the three different conditions by the same scoring system. The response to each statement was scored from 1 to 6, depending upon the degree of agreement indicated by the subject, with complete disagreement receiving a score of 1, and complete agreement receiving a score of 6.

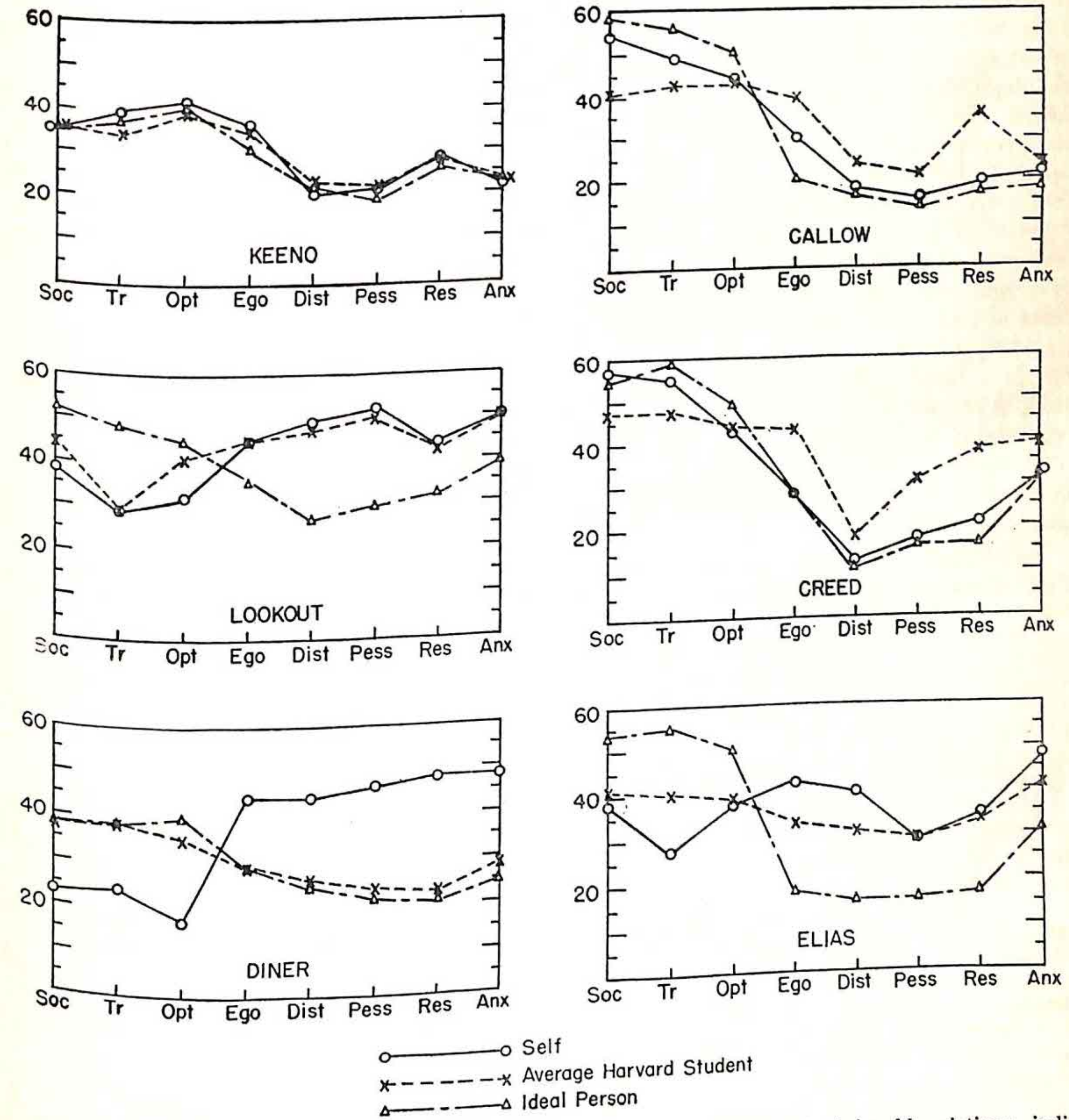

Fig. 1. Affect Questionnaire scores under three conditions. Dispositions, and their abbreviations, indicated on the abscissas are as follows: sociocentricity (Soc); trust (Tr); optimism (Opt); egocentricity (Ego); distrust (Dist); pessimism (Pess); resentment (Res); anxiety (Anx).

Scores assigned to the ten items representative of each of the dispositions were then summed, resulting in a total score for each separate disposition.

The statistical finding of highly significant intercorrelations between these scores, presented in Table 1, led to the formulation of a syndrome for which the term *alienation* seemed most fitting. This syndrome is composed of the following five interrelated dispositions: egocentricity, distrust, pessimism, anxiety, and resentment. Our findings indicate that subjects who are high on any one of these dispositions tend to be high on all of them. The subjects were rank ordered in terms of: (*a*) the amount of alienation evidenced in their self-avowals of agreement with the statements indicative of negative dispositions; (*b*) the amount of alienation they believe would be evidenced by the average Harvard student; and (*c*) the amount of alienation they indicate would be characteristic of their ideal person.

An individual graph was plotted for each subject, showing his scores for each disposition under the three different instructions. A few of the graphs, identified by the subjects' pseudonyms, are shown in Fig. 1. From these scores it was possible to rank order the subjects in terms of the following variables: (*a*) inaccuracy of social apperception; (*b*) discrepancy between self and the apperceived average Harvard student; and (*c*) discrepancy between self and the ideal person.

Profile analysis. Visual inspection of the graphs presented in Fig. 1 indicates marked individual differences. By way of example we will discuss briefly a few of these differences.

Lookout, who is the highest ranking subject on self-alienation, reveals that he apperceives the average Harvard student as only slightly lower on alienation than himself, and that he feels both he and the average student are considerably more alienated than the ideal person. Diner, who is the third highest on self-alienation, apperceives very little discrepancy between scores for the average Harvard student and the ideal person, but shows pronounced discrepancy between these scores and his self scores. Keeno, who ranks seventeenth on self-alienation, shows hardly any discernible difference among his scores under the three different conditions. Although in the present report the focus is on hypothesis testing, rather than diagnosis of individual personalities, it is encouraging to note that the impressions gained from inspection of these profiles are, in most cases, in keeping with diagnoses based on intensive analysis of data secured by means of a variety of standard clinical procedures. At this point, we wish merely to point to the interpretive possibilities of the individual profiles, and suggest that eventually this procedure may prove to be an efficient and fruitful method for assessing personality.

Ego Structure

Before being accepted as a research subject, each individual underwent a brief fact-finding interview with an experienced clinical psychologist. At a later date, after reading the detailed autobiography written by each subject, this psychologist conducted a second personal interview. His clinical evaluations, based on the findings from the autobiography and the two interviews, were expressed by rank ordering the subjects in terms of "ego structure." This concept of "ego structure" or "ego strength" has been defined in detail by Murray and Kluckhohn (11). There is a good deal of clinical evidence of a high degree of interdependence between ego structure, as therein defined, and good personal and social adjustment. In the present study, ego structure has been taken as an indirect measure of adequacy of adjustment.

Hypotheses, Findings, and Discussion

The highly significant intercorrelations between subjects' self scores on the five negative dispositions, which comprise the syndrome termed alienation, led to the formulation of several hypotheses pertaining to personal apperceptions and social apperceptions of people who are high on alienation.

Hypothesis I: Subjects who are high on the syndrome of alienation will be less accurate in their social apperceptions than subjects who are low on the syndrome of alienation.

In order to test this hypothess it was necessary to secure an index of accuracy of social apperception. One criterion of accurate social apperception, most clinicians would agree, is the ability to make accurate judgments in regard to the sentiments and opinions of one's fellowmen. If one is to have satisfactory interpersonal relations, he must be sensitive to the feelings and attitudes of the people in his environment. The more one distorts, or misperceives, the personalities of those around him, the less likely he is to make a good adjustment to his surroundings.

A measure of accuracy of social apperception would be provided by comparing the subject's apperceived "average Harvard student's" responses with the actual average Harvard student's responses. To arrive at scores representative of the mythical average Harvard student's actual responses, the scores obtained by our subjects under the condition of their own self-agreement or disagreement with the statements were averaged. Naturally, our sample is small, and though selection procedures were designed to secure a heterogeneous group, it is possible that the average scores for this group are actually not close to the "true" average of the Harvard college population. Therefore the results presented here may well not represent the final word on this matter. But, with this qualification in mind, the discrepancy between each subject's estimate of the responses of the average person and the actual responses of the average person was taken as a measure of accuracy of social apperception. That is, the less this discrepancy, the more accurate the subject's social apperception.

To test Hypothesis I, the subjects were ranked on the syndrome of alienation, and ranked according to the degree of discrepancy between their estimates of the average and the actual obtained average. The rank-order correlation coefficient (tau = + .142) is in the predicted direction, but does not attain the .05 level of significance. All statistical findings reported in this section are shown in Table 2.

The failure of this hypothesis to be confirmed led to the formulation of the following hypothesis:

Hypothesis II: Subjects who are low on alienation will apperceive the average Harvard student as low on alienation, and those who are themselves high on alienation will apperceive the average Harvard student as high on alienation.

Rank-order correlation between the subjects' own rank on alienation and their rank on the basis of the amount of alienation apperceived in the average person is significant at the .003 level (tau = + .453), which confirms Hypothesis II.

Table 2

Self-Alienation on the Affect Questionnaire Correlated with Other Measures on the Affect Questionnaire, and with Ego Structure Based on Clinical Evaluation
($N = 20$)

Measure	Correlation (tau [8]) with self-alienation	Level of significance*
Alienation of apperceived "average Harvard student"	+.453	.003
Discrepancy between self and apperceived "average Harvard student"	+.429	.005
Discrepancy between self and the ideal person	+.516	.001
Inaccuracy of social apperception	+.142	.142
Ego structure	−.411	.006

* One-tailed tests of significance.

These findings suggest that there are inaccuracies in the social apperceptions of both low-alienation subjects and high-alienation subjects, with a trend in the direction of a higher degree of accuracy in those subjects who are low on alienation. Although there is not a significant association between alienation and accuracy of social apperception, confirmation of Hypothesis II indicates that the type of apperceptual error is significantly different in lows and highs, with subjects who are themselves low on alienation perceiving other people in their environment as low on alienation, and those who are high on alienation perceiving others as high on alienation. In other words, defining "better" and "worse" in terms of cultural sanction and social acceptability, low-alienation subjects see people as better than they really are, and high-alienation subjects see people as worse than they really are.

This phenomenon provides a vivid example of the operation of self (supplementary) projection, which is the tendency for people to perceive in others those things that are actually in themselves. In classic Freudian terminology the concept of projection was employed solely to describe the attribution of one's negative, undesirable traits on to other people. But the present results suggest that the concept of projection is applicable to positive attributes as well as negative attributes, with happy, contented, well-adjusted individuals tending to perceive others as similar to themselves, and, consequently, seeing them as less alienated than they really are. Those who are low on alienation look around them and see a lot of trustful, happy, sociocentric human beings, whereas those who are high on alienation look at these same people and decide that they are self-centered, resentful, anxious, and unhappy.

In the attempt to discover additional relations between alienation and social apperception the following hypothesis was tested:

Hypothesis III: The degree of negative discrepancy between the subject's own responses and the responses he attributes to the average Harvard student will be greater in subjects who are high on alienation than in subjects who are low on alienation.

There are two ways to receive a score indicative of negative discrepancy. The first is to apperceive the average Harvard student as being higher on a positively evaluated personality disposition than is the self, and the second is to apperceive the average Harvard student as lower on a negatively evaluated personality disposition than is the self.

Correlating rank on amount of negative discrepancy between self and the apperceived average Harvard student with rank on self-alienation results in a tau of + .429, which is significant at the .005 level. This finding indicates that individuals who are at either extreme on the variable of alienation tend to rate others as closer to the mean than themselves.

There remains one crucial relationship within the Affect Questionnaire, the comparison of self scores on alienation with alienation scores of one's ideal person.

The following hypothesis is concerned with this relationship:

Hypothesis IV: The degree of negative discrepancy between the subject's own responses and the responses of his ideal person will be greater in subjects who are high on alienation than in subjects who are low on alienation.

To test this hypothesis, rank on amount of negative discrepancy between self and the ideal person was correlated with rank on self-alienation. The resulting coefficient (tau = + .516) is highly significant, and confirms the above hypothesis. This finding confirms the widely held belief that an important correlate of unhappiness, and probably maladjustment, is the perception of the self as falling below one's standards and ideals.

As alienation is believed to be a sign of emotional maladjustment, it seemed desirable to correlate rank on self-alienation with some independent measure of adequacy of personal and social adjustment. The measure selected for this purpose is the psychologist's rank order of the subjects in terms of ego structure. With these two measures, the following hypothesis was tested:

Hypothesis V: Subjects who are low on ego structure will be high on the syndrome of alienation.

The predicted negative association was found (tau = − .411, $p = .006$) indicating that subjects who view the world and their relationship to it in a negative light have weak ego structure (poor adjustment) as evaluated by experienced clinical judgment. This finding of highly significant agreement between results of assessment by these different methods speaks for the validity of the Affect Questionnaire as an objective method for investigating personality and apperception.

Conclusions

The over-all findings suggest the following picture of an alienated young man: He is a person whose apperceptions reveal relatively large amounts of egocentricity, distrust, pessimism, anxiety, and resentment in his personality. His social apperception is distorted in that he apperceives other people in his environment as being more alienated than they actually are. Although he sees others as alienated, he apperceives his contemporaries as less alienated than himself, consequently less miserable than himself. And, finally, he feels that his negative evaluations of the world and its people are a far cry from the evaluations that would be characteristic of any person he would regard as admirable or ideal.

The testing of the above hypotheses has revealed that in our sample there are individuals who see themselves as far from their ideals and considerably worse off, socially and psychologically, than their contemporaries. This syndrome of alienation, shown to correlate with weak ego structure, which permeates their self-picture and distorts their apperceptions of others, must certainly influence their cognitive processes and their social relations. The cognitive and behavioral consequences of this syndrome should provide fertile ground for future investigation.

Summary

The purpose of this study was to develop an objective instrument for measuring apperception, and to test a series of hypotheses concerning relations between personal apperception, social apperception, and ego structure. Twenty Harvard undergraduates responded to a questionnaire containing evaluative statements representative of certain critical dispositions of personality. Each subject responded to these statements under the following conditions or sets: (*a*) his own personal agreement or disagreement with them; (*b*) the degree of agreement he thinks would be indicated by "the average Harvard student"; and (*c*) the degree of agreement he thinks would be indicated by an ideal person.

Statistical analyses of results on this questionnaire led to formulation of a syndrome, termed *alienation,* composed of the following five interrelated dispositions: egocentricity, distrust, pessimism, anxiety, and resentment. Our findings indicate that alienated subjects are "lone wolves" with grievances, distrustful of their fellowmen, apprehensive and gloomy in their anticipations of the future.

Empirical examination of five hypotheses revealed that, in our sample, individuals who are high on this syndrome tend to apperceive other people in their social environment as more alienated than they really are, yet less alienated than themselves. Also, individuals who are high on alienation showed a greater discrepancy between themselves and the apperceived average person and between themselves and their ideal person than did individuals who are low on alienation. Finally, it was found that individuals who view the world and their relationship to it in a negative light have a weak ego structure as estimated by experienced clinical judgment.

Received May 24, 1954.

References

1. Bender, I. E., & Hastorf, A. H. The perception of persons: forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.,* 1950, **45**, 556–561.
2. Bruner, J. S., & Tagiuri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 634–654.
3. Chowdhry, Kalma, & Newcomb, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.,* 1952, **47**, 51–57.
4. Fensterheim, H., & Tresselt, Margaret E. The influence of value systems on the perception of people. *J. abnorm. soc. Psychol.,* 1953, **48**, 93–98.
5. Gage, N. L. Accuracy of social perception and effectiveness in interpersonal relations. *J. Pers.,* 1953, *22*, 128–141.
6. Hites, R. W., & Campbell, D. T. A test of the ability of fraternity leaders to etimate group opinion. *J. soc. Psychol.,* 1950, **32**, 95–100.
7. Ichheiser, G. Misunderstandings in human relations: a study in false social perception. *Amer. J. Sociol.,* 1949, **55**, Part 2.
8. Kendall, M. G. *Rank correlation methods.* London: Griffin, 1948.
9. Murray, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.,* 1933, **4**, 310–329.
10. Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
11. Murray, H. A., & Kluckhohn, C. Outline of a conception of personality. In C. Kluckhohn, H. A. Murray, & D. M. Schneider (Eds.), *Personality in nature, society, and culture.* New York: Knopf, 1953. Pp. 3–49.
12. Notcutt, B., & Silva, A. L. M. Knowledge of other people. *J. abnorm. soc. Psychol.,* 1951, **46**, 30–37.
13. Soskin, W. F. Influence of information on bias in social perception. *J. Pers.,* 1953, **22**, 118–127.
14. Stevenson, B. *The home book of quotations.* New York: Dodd, Mead & Co., 1934.
15. Tagiuri, R. Relational analysis: an extension of sociometric method with emphasis upon social perception. *Sociometry,* 1952, **15**, 91–104.
16. Tagiuri, R., Blake, R. R., & Bruner, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.,* 1953, **48**, 585–592.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Relationships Between the Rosenzweig P-F Study and Test Duration, Socioeconomic Status, and Religion[1]

Edward T. Parsons
Pennsylvania State University

This study had four aims: to correlate Rosenzweig Picture-Frustration Study scores with the time taken in answering it; to correlate P-F scores with scores of the Sims Score Card for Socio-Economic Status; to investigate differences between P-F scores of Protestants and Catholics as a group and in majority and minority situations; to compare P-F scores of those omitting items of the P-F study to P-F scores of the total group, and determine the P-F items most often omitted and unscorable.

The *S*s were 150 white seniors from four Pennsylvania high schools. Each school supplied a whole class, a class in a course that students were required to take, and a class to which they had been assigned in a random fashion.

Sims scales were administered first, then the P-F study. The time each student used to answer the P-F was recorded. Product-moment coefficients of correlation were computed to investigate relationships between P-F scores and test duration and between P-F scores and socioeconomic status scores.

Need persistence and intropunitiveness correlated positively with test duration, both .24 ($p = .01$). Extrapunitiveness correlated negatively with test duration, − .18 ($p = .05$).

Socioeconomic status scores ranged from 2 to 31, "lowest" to "highest," according to Sims' provisional levels. The mean was 17, "high." No significant correlation was found between socioeconomic status scores and P-F scores.

Catholic/Protestant percentage ratios were 27/73 and 75/25 in two classes. P-F mean scores of these groups were compared to investigate differences between Catholics and Protestants in minority and majority situations. One difference was significant at the .05 level, but chance would account for this since 28 differences were checked. Comparing P-F mean scores of all Catholics to scores of all Protestants gave no significant differences.

P-F mean scores of the 13 *S*s who omitted one or more P-F responses were compared to mean scores of the total sample. Those who omitted responses earned higher mean group conformity ratings. The difference was significant at the .05 level.

The number of omitted and unscorable responses was less than 2% of all responses. Item 19 of the P-F study was most often omitted, and 69% of all omissions occurred in the second twelve items. Unscorable responses were found most often with items 14 and 20, and 75% of the unscorable responses were in the second half of the P-F study.

Brief Report
Received October 26, 1954.

---

[1] This report is based upon an M.S. thesis supervised by Dr. William U. Snyder at the Pennsylvania State University.

An extended report may be obtained without charge from E. T. Parsons, Children's Service Center, 335 S. Franklin St., Wilkes-Barre, Pa., or for a fee from the American Documentation Institute. To obtain it from the latter, order Document No. 4455 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Organizational Activity on the Rorschach Test[1]

Boyd D. Sisson
VA Hospital, Omaha, Nebraska

and Earl S. Taulbee
Norfolk State Hospital and University of Nebraska College of Medicine

Organization in Rorschach responses—defined by Beck as the combination of test figures into larger units that may be whole responses, details seen in relation to each other, or the combination of white spaces with filled-in elements—is widely used in the evaluation of test protocols. The meaning commonly attached to this test activity is as an indicator of intelligence or a component of intelligence, with the assumption of a direct and positive relationship between the amount of organization and level of intelligence. The literature, however, contains relatively little research specific to the assumption involved in the common interpretation of the organization factor. This paper will report briefly some hitherto unpublished findings pertinent to this area. It also will provide additional information to a published study by Wilson and Blake (4) concerning a methodological problem in the organizational concept.

In a recent study involving the Rorschach test and the SRA Primary Mental Abilities Test with a group of 32 "normal" high school students, Batt (1) reported correlation coefficients of .49 and .46, both statistically significant at the 1% level, between weighted *Z* score and the Verbal and Reasoning portions of the PMA, respectively. He interpreted these findings as "indicating *Z* to relate most closely to particular kinds of intellectual abilities, specifically those involving capacity to understand ideas expressed in words, to solve logical problems, and to foresee and plan."

In another study Taulbee (3) used the Rorschach to evaluate the intellectual levels of functioning in a group of 60 hospitalized schizophrenic patients. He found the *Z* score to have a positive and significant relationship with the Full-Scale Wechsler IQ, but that the relationship in this group was due almost entirely to the correlation with three of the subtests—Digit Span, Similarities, and Vocabulary. It was concluded from these results that among schizophrenic patients the ability to organize and see things as a whole seems to be related to freedom from distractability, the ability to engage in verbal abstraction, and the ability for concept formation.

Similarly, Wishner (5) previously reported Beck's *Z* to be the most consistently significant of the Rorschach signs as an index of intellectual performance for a group of neurotic patients based on a correlation at the 1% level of confidence between *Z* and the Full-Scale Wechsler IQ. These two studies show a consistency in significant correlations of *W* and *Z* with Vocabulary subtest scores, which offers support for the widely used clinical hypothesis that vocabulary test scores are relatively resistant to the effects of neurotic or schizophrenic processes and, therefore, provide good estimates of premorbid intellectual functioning.

Based on a group of 50 randomly selected psychiatric patients, Sisson and Yager obtained a correlation of .428 ($p < .01$) between weighted *Z* score and Wechsler Full-Scale IQ; their data also yielded a correlation of .452 between unweighted number of *Z* responses and Full-Scale IQ. Thus, it is seen that the above findings are in agreement in supporting the validity of the assumption that there is a positive relationship between or-

---

[1] From the VA Hospital, Omaha, Nebraska.

Table 1

Relationship Between Weighted *Z* Responses and Number of *Z* Responses

Group	*N*	Age range	Mean age	Mean weighted *Z*	*SD*	Mean no. of *Z*	*SD*	*r**
Normal adults	60	20–52	36.0	24.03	18.32	8.45	5.36	.982
Normal children	70	8–18	13.4	23.22	19.40	7.71	5.31	.988
Schizophrenics	60	20–39	27.6	27.38	17.30	9.08	5.17	.976
Neurotics	35	20–46	31.9	26.87	19.59	9.57	6.14	.985
Organics	10	19–44	32.8	27.50	15.11	8.90	4.01	.983

* Product-moment correlations between weighted *Z* scores and number of *Z* responses.

ganization activity on the Rorschach test and intelligence.

As pointed out by Wilson and Blake (4), a second assumption involved in the organizational score aside from its operational validity is the technique of scaling organization. They reported a study concerned with an analysis of the weighting system used by Beck in connection with the various types of organizational activity. Their results showed a "consistent and extremely strong relationship" between simple summation of responses containing *Z* and total weighted *Z* scores. Their data yielded no product-moment correlation coefficient smaller than .98 for a varied group of 104 subjects, from which they conclude that determination of the types of organized responses, the weights for each, or the sum of weighted *Z* values contributes little more than can be had simply by counting the number of organized responses. In the same report they present a table for predicting total weighted *Z* score when number of *Z* responses is known.

The following study is reported for the purpose of providing additional data to those of Wilson and Blake.

Study

The Rorschach protocols of 235 subjects, representing five groups, were analyzed to determine the relationship between the weighted *Z* scores and number of *Z* responses. All subjects were males except for the normal children group which consisted of 29 males and 41 females. The schizophrenic, neurotic, and organic subjects were all hospital patients at the time of testing; the normal subjects were from a small midwestern community[2] located in the same geographical locale from which the patients came.

Table 1 gives a summary of the calculations. Analysis of the data shows that employing the table of Wilson and Blake for best total weighted *Z* prediction when total number of *Z* responses is known (which has a standard error of 4.74) correctly classified the following percentage of cases within one and two standard errors, respectively: 83% and 97% of the normal adults, 90% and 99% of the normal children, 73% and 98% of the schizophrenics, 71% and 91% of the neurotics, and 90% and 100% of the organics.

Discussion

Results of the studies presented above support the assumption of a direct and positive relationship between organization activity on the Rorschach test and intelligence. In this respect, then, the findings contribute to the problem of operational validity of the organization factor. There is additional suggestion that at least with certain groups the relationship between *Z* and specific components of intelligence is greater than it is with the total or generalized intelligence functioning. This is viewed as denoting the areas of functioning which seem to be least affected by the schizophrenic process, but it may also be helpful in generating hypotheses as to the specific processes involved in the ability—or inability—to organize one's percepts.

---

[2] The records were from a larger study done under the supervision of Dr. Marshall R. Jones and supported by the University of Nebraska Research Council.

It is the writers' opinion that the findings concerning the relationship between weighted Z scores and number of Z responses show that quantitatively the present system of scaling the Z responses for the purpose of enhancing the differentiating function of this activity is of little value.

When the organization factor is used as a single total as its contribution to interpretation of the Rorschach test, the findings have significance not only in questioning the value of the weighting system, but perhaps also in the interest of time saving and simplicity. It is recognized, however, that the use of Z in interpretation is not necessarily confined to mere consideration of its total amount, but that it probably offers much more when also utilized by analyzing its kind, distribution, or absence. In other words, the writers feel that the rationale for scaling Z according to the nature of the organizing activity involved has merit, and they certainly do not advocate abandoning it but believe that the above findings show the necessity for further research in establishing a more sensitive method of scaling.

Summary

Results of several unpublished studies are reported concerning the relationship between the organization factor on the Rorschach test and intelligence level. In the main, they support the clinical hypothesis of a direct and positive relationship, and also present evidence that the Z activity, at least among certain groups, seems to be related most closely with certain of the verbal functions in the intelligence concept.

Report is made of a study which yields very close agreement with earlier findings by Wilson and Blake of a consistent and extremely strong relationship between simple summation of responses containing Z and total weighted Z scores.

Received May 12, 1954.

References

1. Batt, H. V. An investigation of the significance of the Rorschach Z score. Unpublished doctor's dissertation, Univer. of Nebraska, 1953.
2. Beck, S. J. *Rorschach's test.* Vol. 1. *Basic processes.* (2nd Ed.) New York: Grune & Stratton, 1949.
3. Taulbee, E. S. The value of the Rorschach test for evaluating the intellectual levels of functioning in schizophrenics. Unpublished doctor's dissertation, Univer. of Nebraska, 1953.
4. Wilson, G. P. Jr., & Blake, R. R. A methodological problem in Beck's organizational concept. *J. consult. Psychol.*, 1950, **14**, 20–24.
5. Wishner, J. Rorschach intellectual indicators in neurotics. *Amer. J. Orthopsychiat.*, 1948, **18**, 265–279.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Psychiatric Screening of Flying Personnel with the Cornell Word Form[1]

John R. Barry, Saul B. Sells, and David K. Trites
Air Force School of Aviation Medicine, Randolph Field, Texas

The authors elsewhere (1) have summarized research in progress on the psychiatric selection of flying personnel. The present report describes the development and validation for psychiatric screening of a key for the Cornell Word Form (CWF), a controlled word-association test. This test was one of twelve experimental personality tests administered to 1,036 cadets entering pilot training at Randolph Air Force Base in 1949 and 1950. None of these cadets was excluded from training on the basis of his test responses.

Three scoring keys were selected for validation. Subsequently, the most efficient key was cross-validated with significant results in three independently selected, but overlapping, samples, selected to define different, but related, criteria of adjustment. In the first cross-validation sample, cadets who resigned from training were considered the lower criterion group members ($N = 128$); cadets who completed primary flying training ($N = 301$) or who were eliminated because of flying inaptitude ($N = 210$) were included in the upper criterion group. When the key was applied to the test responses of these three groups, 14 per cent of the resignee group (true negative) and less than 8 per cent of the other two groups (false negative) were identified as potentially unfit. A significantly larger proportion of the resignees than of the other two groups was identified ($p < .03$).

In a second study, criterion status was based upon psychologists' over-all ratings of adjustment to flying training rather than upon training outcome alone. Cadets who failed and who were given low ratings were assigned to the lower criterion group ($N = 309$). The upper criterion group ($N = 392$) included cadets who were judged to have adjusted to flying better than this lower criterion group. As was predicted, a significantly larger proportion of the lower than of the upper criterion group (11 and 7 per cent respectively) made three or more unhealthy test responses ($p < .04$).

In a third study, a stratified sample was selected to represent the total population with reference to the adjustment ratings and pass-fail status. However, only the ratings were used in the criterion. When these ratings were dichotomized at their mean, a significantly larger proportion of those below than of those above the mean (13 and 4 per cent respectively) was identified by CWF responses ($p < .02$).

It appeared from this research that a small but significant number of the cadets who adjust poorly to flying training can be identified by their responses to the CWF. In these responses the cadets seem to express certain fears, dissatisfactions, and psychosomatic symptoms.

Brief Report
Received October 22, 1954.

Reference

1. Sells, S. B., & Barry, J. R. A research program to develop psychiatric selection of flying personnel. *J. aviat. Med.*, 1953, 24, 29–47.

---

[1] An extended report of this study may be obtained without charge from John R. Barry, USAF School of Aviation Medicine, Randolph Field, Texas, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4452 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.50 for microfilm or $6.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Threshold for the Perception of Human Movement in Inkblots[1]

Frank Barron

Institute of Personality Assessment and Research, University of California, Berkeley

One of the variables conceptualized and scored as part of the Rorschach Psychodiagnostic (7) is the tendency of the subject to perceive human beings, or human-like animals, engaged in characteristically human activity in the blots (the *M* response). This response tendency is said to go along with a bent toward introversion rather than extraversion. Introversion has been variously defined; in general, it may be conceived of as a tendency to take thought rather than action, to experience vicariously and in fantasy rather than in real life and directly, and to create through the arrangement of symbols rather than through the arrangement of objects which stand for nothing but themselves.

The aim of the present investigation was to study the personality correlates of the human movement response tendency in a research situation which would make it possible to obtain independent evidence concerning the personalities of the subjects. Such a situation is provided by the living-in assessment method, with its emphasis upon observation of the subjects through several days of informal social interaction, lifelike situational tests, group discussions, improvisations, group games, and the like. This makes it possible to discover how persons who have much or little readiness to perceive human movement in inkblots are described by skilled psychological observers who become acquainted with them but not with their test scores.

The measure of readiness to perceive human movement in inkblots was especially constructed for this study, and represents an attempt to supplement the Rorschach test itself as an instrument of observation when Rorschach theory is in question. As users of the test know, the Psychodiagnostic is very complex, both stimulus-wise and in its scoring scheme, and it does not lend itself well to any attempt to isolate variables and to separate out their correlational components. While this complexity is important to the test as a vehicle for clinical observation, it contributes to certain psychometric shortcomings and unnecessary difficulties when the verification of theory is the chief concern.

One important difficulty with the scoring scheme is that the number of responses varies widely for different subjects. Productivity is itself an important variable, of course, but the present method of obtaining a measure of it tends to confound the evaluation of other measures which may be equally important. Subjects now cannot be compared in terms of absolute incidence of a given type of response, since this is partly a function of total number of responses. Furthermore, subjects cannot be compared in terms of *relative* incidence, for relative incidence of response in a given category varies in some nonlinear and as yet undetermined fashion with total number of responses. This in turn is dependent to some extent on stimulus properties of the blots; there is clearly a limitation to the number of responses which can be given in any single category, and after a certain point in the production of responses the more limited categories begin to suffer relative to the others.

What is needed, then, is a method of keeping the number of responses more or less constant for all subjects, while yet providing considerable opportunity for the subject's response tendencies to emerge. At the same time, stimulus strength should be weighted properly in evaluating response strength; one difficulty with the Rorschach measure of *M* tendency is that it is a simple count of the number of human movement percepts which are verbalized by the subject, without regard for the power of the stimulus to evoke a human movement response in the average person.

---

[1] This research is supported in part by the United States Air Force under Contract No. AF 18 (600)–8, monitored by Technical Director, Detachment No. 7 (Officer Education Research Laboratory), Air Force Personnel and Training Research Center, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government. Personal views or opinions expressed or implied in this publication are not to be construed as necessarily carrying the official sanction of the Department of the Air Force or of the Air Research and Development Command.

Some of the difficulties may be met if these prescriptions are followed: (*a*) increase the number of blots; (*b*) score only one response, the first, to each blot; (*c*) take systematic account of the relationship between stimulus strength and response tendency by employing the conventional experimental index of this relationship, response threshold; (*d*) isolate the main Rorschach variables and study them one at a time before attempting to study them all together.

The rationale of these prescriptions is simple and clear. An increase in the number of blots should achieve more representativeness on the stimulus side and more total-score reliability (since reliability may be increased, up to a point, by increasing test length). Scoring of only one response to each blot makes the absolute number of responses in each scoring category comparable from subject to subject, and makes feasible the use of some sort of standard score, such as Z scores, so that the individual subject's performance may be immediately referred to that of the general population.[2] The weighting of stimulus strength in evaluating response strength is essentially a more differentiated way of scoring, comparable to the use of refined rather than crude weights in prediction; the addition of the concept of threshold makes the human movement phenomenon more assimilable to established knowledge and methods in experimental psychology, which is all to the good so far as Rorschach theory is concerned. Finally, the study of variables in isolation, however unholistic it may seem, may really be the best possible foundation for the understanding of variables in interaction.

In the present study, then, the variable under consideration is threshold for perception of human movement in inkblots; the aim of the study is to construct a measure of threshold and to ascertain its personality correlates; the general method of personality research employed is the living-in assessment method.

Construction of the Measure of Threshold

The model for the construction of a measure of threshold for perception of human movement is the conventional stimulus series used to determine response thresholds in such sense modalities as the auditory, olfactory, tactile, and the like. Although stimulus strength or intensity is not determinable from physical properties of the stimulus in the case of inkblots, this is no great loss so long as relative frequencies of response can be established in large samples and with some stability. By arranging inkblots of known relative frequency in a regularly graduated series, with p values ranging from .00 to 1.00, a measure analogous to the usual perceptual stimulus series is constructed. The subject's threshold for human movement is then the ordinal position of that blot in the series at which he first gives a human movement response.

This ideal design was not carried out exactly as projected in the present study, largely because of time pressures in the scheduling of the assessments. One hundred and fifty achromatic inkblots had been constructed, using 4- by 6-inch sheets of white paper, which were then mounted on stiff cardboard of the same dimensions. The plan had been to select 26 blots from the 150, on the basis of observed frequencies, in such a manner as to make a series with graduations of approximately .04. It became necessary, however, to select the blots on the basis of observations in a small preliminary sample, and to order them initially in terms of observed frequencies in that sample. For this reason, the threshold score finally assigned to the assessed subjects was based on the relative frequency of M response to each blot *in the assessed sample itself*. This design is somewhat inferior to what was first intended, but it still fills the prescriptions specified for construction of a measure which would overcome some of the difficulties inherent in the Rorschach measure of M tendency.

As it turned out, the preliminary standardization yielded a fairly stable stimulus series. The rank-difference correlation between the order of presentation based on frequencies in the preliminary standardization group and the final ordering of the blots in terms of frequencies observed in the assessed sample is .85. The stability of the final ordering of the blots in the rather homogeneous Air Force officer sample is indicated by the rank-difference correlation between a rank order based on observed frequencies in the first 50 cases and a rank order based on observed frequencies in the final 50 cases: rho is .99. For this homogeneous sample, the *M*-evocative power of inkblots is remarkably stable.

---

[2] It might be pointed out incidentally that Rorschach ratios, such as M:Sum C, may be much more meaningful if they are ratios of Z scores rather than absolute scores; Z ratios would be free of the often unrecognized and cumbersome assumption underlying absolute score ratios, viz., that the blots themselves present precisely equal opportunity for the two contrasting experience-types, introversive and extratensive, to manifest themselves.

The possible range of scores for individual subjects is, of course, from 1 to 27 (the score of 27 being assigned to subjects who do not see *M* in any one of the 26 blots). The expected range, considering the method of construction of the measure, is from 2 (since no subject should see *M* on the first blot) to 26. The actual observed range was from 2 to 25, with the mean at 13. The form of the distribution was approximately rectangular.

The Sample and the Method of Study

The subjects in this study were 100 captains in the U. S. Air Force. As a group they were above average in intelligence, in education, in physical health, and in personal stability. The age range was from 27 to 50, with a mean age of 33. All of the subjects were men. All but three were married, and most of them had at least two children. In pre-army socioeconomic background they tended to be lower middle class. The majority of these officers were combat veterans, and many of them had been decorated for valor in World War II. In most ability measures they scored well above average, and were less variable than men-in-general.

The subjects were seen in groups of ten for three full days at the Institute house. The method of study placed considerable emphasis upon observation of the subjects in informal social interaction. The assessment staff psychologists sat down to meals with the subjects, took part in a social hour before dinner, and in general were participant observers throughout the three days of living-in assessment. Situational tests, interviews, group discussions, improvisations, charades, and the like were included in the assessment program, so that the social characteristics of the subjects had much opportunity to manifest themselves, and the raters were in a position to observe significant behavior.

Observations made by staff members during the three days of study were condensed and rendered easily susceptible of statistical analysis chiefly through two techniques: a set of 76 statements descriptive of personal functioning and permitting the expression of clinical inferences, and an adjective check list (3) consisting of common, personally descriptive adjectives. Both techniques were used by staff members at the conclusion of the three days of assessment to sum up their impressions of each subject.

The 76 statements were sorted on a 9-point scale (a so-called *Q* sort), the frequencies at each point being such as to make the final distribution conform closely to the normal curve. The 300 adjectives were checked simply as characteristic or not characteristic of the given subject. The *Q*-sort composite was obtained by averaging the placements of the items by four staff members, and then redistributing these averaged values on a normal curve. The adjective composite consisted of all adjectives which had been checked as characteristic of a subject by three or more raters.

These descriptions and ratings were of course given without knowledge of the objective test performances of the subjects. No rater knew the *M* threshold of any of the subjects at the time the ratings were made.

In addition to these descriptions based on social observation of the subjects, information is available concerning life-history factors. Each subject was interviewed for two hours by a psychiatrist or clinical psychologist (who did not otherwise participate in the assessment), and again a check-list method was used to summarize impressions from the interview. The life-history check list contained sections having to do with the subject's recollection of himself as a child, his imagery of his parents, his favorite games, his playmates, and so on.

Finally, certain objective test measures were obtained on all subjects. These included measures of general intelligence, of originality, of personal stability, and of social attitudes. Scores from all of them were available for correlation with *M* threshold.

Correlates of *M* Threshold

It should be remembered, of course, that high scorers on the threshold measure are those with less readiness to perceive human movement; lower threshold means greater disposition to give an *M* response.

The observer adjective descriptions which are significantly (.05 level) related to threshold for *M* when the first quartile of the distribution (threshold scores from 2 to 8) is compared with the fourth quartile (threshold scores from 18 to 25) are as follows:

Early M *response (low threshold)*

intelligent
fair-minded
mild
anxious
mannerly
inventive
interests wide

Late M *response (high threshold)*

practical
stubborn
simple
masculine
arrogant

The composite *Q*-sort descriptions yielded the following items with significant (.05 level) differences between the groups:

Early M *response (low threshold)*

1. Highly cathects intellectual activity; values cognitive pursuits.
2. Gets along well in the world as it is; is socially appropriate in his behavior.
3. Is introspective; concerned with his self as object; frequently self-aware.
4. Has high degree of intellectual ability.

Late M *response (high threshold)*

1. Has narrow range of interests.
2. Allows personal bias, spite, or dogmatism to enter into his judgment of issues.
3. Prefers action to contemplation.
4. Is rigid; inflexible in thought and action.

At this point, an important discrepancy between observers' descriptions and actual test measures should be noted. Although subjects with greater *M* tendency were described as inventive, intelligent, and broader in their interests, correlations between *M* threshold and actual measures of intelligence, general information, and originality in problem solving were all in the neighborhood of zero. The aptitude tests used included the Terman Concept Mastery Test,[3] the Wesman Personnel Classification Test (9), the Arthur Stencil Design Test (1), the Idea Classification Test,[4] and the Test of Mechanical Comprehension (2). A 180-item General Information Survey constructed by Gough (4) was also employed. Several tests from the Guilford Creativity battery (6), including the Unusual Uses, Plot Titles, Consequences, Match Problems, Gestalt Transformation, and Controlled Association tests, were used; none showed a significant relationship to *M* threshold. The failure to find such relationships occurred both when *M* threshold was correlated with these measures for the entire 100 cases, and also when the first quartile on *M* was compared with the fourth quartile.

What these facts plainly say is that subjects with greater readiness to perceive *M* appear to psychological observers to be more intelligent, more inventive, and broader in their interests in social situations, but that in controlled testing situations, when problem-solving activity is called for, *M* threshold is unrelated to test scores. If we assume the validity of the test measures, the most direct explanation of the discrepant adjective and *Q*-sort descriptions would seem to be that psychologists falsely attribute greater intelligence to persons who manifest more tendency to introspect, to take thought rather than to take action. A check on this hypothesis seems to support it: the *Q*-sort item "Has a high degree of intellectual ability" correlates .45 with the item "is introspective; concerned with his self as object; frequently self-aware." The latter item, however, correlates only .11 with the Arthur Stencil Design Test, .02 with the Idea Classification Test, .26 with the Wesman Personnel Classification Test, and .27 with the Terman Concept Mastery Test. Introspection thus appears to be weighted unduly by the psychological staff in judging intelligence. Perhaps, after all, the more "thoughtful" person is not always the more intelligent person, thoughtfulness being an intellectual disposition rather than an intellectual ability.

Some evidence concerning developmental factors related to *M* tendency is furnished by the life-history interview material. The subjects with low thresholds for *M* described themselves significantly (.05 level) more often as *quiet* and *self-conscious* during their childhoods. They were also *less interested in sports*, and took *less part in rough-and-tumble play*, than high-threshold subjects; the latter more often described themselves as having been *active*, and characterized their childhood as *happy*.

---

[3] For a general description of the form of this test, which is as yet unpublished, see Terman and Oden (8). The test was made available through the generosity of Dr. Terman.

[4] Acknowledgment is made to Educational Testing Service for making this test available prior to its publication.

There are some possibly related differences between the two groups of subjects in their descriptions of their mothers: the adjectives showing differences significant at the .05 level are *nervous, protective, conservative, punishing,* and *worrying,* all of them being more characteristic of the mothers of low-threshold subjects. The general picture we get from the life-history interview material is that the person who in adulthood has more disposition to see human movement in the inkblots was in childhood somewhat shy, sensitive, and overprotected, or even maternally dominated. In relation to this, it is of interest that the Femininity key of the California Psychological Inventory (5) correlates − .26 with *M* threshold (the more *M*-disposed subjects obtaining higher femininity scores).

In brief, readiness to see human movement in the inkblots seems to go along with *introspectiveness, thoughtfulness, fair-mindedness* (suggesting ethical discrimination based on cognitive principles), and *a high valuation of the intellectual processes,* based perhaps upon a certain inhibition of activity and greater recourse to fantasy in childhood. Such persons are seen in adulthood as relatively *mild, mannerly,* and even *anxious* as compared with the *practical, masculine,* and *simple* individuals who have less tendency to see movement in the blots. It appears too that an extreme lack of *M* tendency may at times be nonadaptive: the *Q*-sort items "is rigid; inflexible in thought and action," and "allows personal bias, spite, or dogmatism to enter into his judgment of issues" point clearly to the kinds of liabilities which accompany the static perceptual attitude. This is seen as issuing on occasion in *stubbornness* and *arrogance.*

Discussion

Rorschach's claim that movement responses are produced most abundantly by persons who "function more in the intellectual sphere, whose interests gravitate more towards their intrapsychic living rather than towards the world outside themselves" (7, p. 64), seems in general to be supported by these findings. In a sense, the fact that *M* threshold is unrelated to measured intelligence strengthens rather than weakens this conclusion. If two groups of subjects are equal in intelligence, but one group is characteristically described as more intelligent, more introspective, more concerned with intellectual activity, and more given to contemplation than to action, then it seems fair to say that these latter individuals "function *more*" (not necessarily better) in the "intellectual sphere."

Rorschach himself, of course, thought of this tendency as "a component of intelligence" (7, p. 63). The production of many movement responses is, he claims, one of the characteristics of "intelligent subjects." Since this observation was based upon his own estimates of his subjects' intelligence, it may be that he was misled by the same occupational prejudice displayed by the psychologists of the staff of this Institute. (Rorschach was not above injecting his own biases rather freely into his experimental observations, as witness his entertaining preoccupation with "pedants" and "grumblers," two diagnostic categories which appear more frequently in the Psychodiagnostic than would seem warranted by ordinary nosological usage.)

In some ways, it seems preferable, from a psychological assessment standpoint, to have a well-defined and measurable variable which is entirely stylistic in character and free of correlation with measures of general ability. Where style is important in effectiveness of performance, such variables may add more to a prediction equation than do variables which are loaded with intelligence factors. This is especially true where personal interaction is concerned, and where certain combinations of styles of performance work better than others. Two-person situations with subjects selected especially for similarities or differences of style offer interesting possibilities for tests of hypotheses concerning the conditions which make for constructive and empathic interaction versus those which make for destructive, stultifying relationships. Rorschach suggests, for example, that a person with strong introversive tendencies can hardly be understood at all by a person whose bent is strongly extratensive; with adequate measures of such tendencies in perception, experimental verification of such a hypothesis would not be difficult.

The present measure of *M* threshold is as yet in a preliminary stage of development. It is evident that Guttman scale analysis is made to order for the psychometric goal of accurate placement of an individual in relation to others on such a dimension as *M* tendency.

It seems quite possible that other Rorschach factors could be equally well measured, and that a series of separate tests could be produced which would be much better, from the point of view of perception theory and research, than the instrument with which Rorschach made his initial observations. He himself, of course, was quite open to such innovations, and indeed suggested a number of interesting possibilities in his classical monograph.

Summary

A measure of threshold for the human movement response was constructed by arranging inkblots of known *M*-evocative power in a regularly graduated series. The final series consisted of 26 blots, selected from a group of 150 blots for which the relative frequency of *M* response was known. The subject's threshold score, based on the final arrangement of the 26 blots, was given by the ordinal position of that blot in the series at which he first verbalized a human movement response.

The correlates of *M* threshold were studied in a sample of 100 military officers who took part in three days of living-in assessment, and who had been described by staff psychologists through the use of *Q*-sort and adjective check-list techniques. The psychologists did not, of course, know the *M* threshold scores of the subjects. It was found that *M* threshold was uncorrelated with measures of intelligence, originality, and associational fluency, but that subjects who displayed considerable readiness to give human movement responses were considered by staff psychologists to be more intelligent, inventive, introspective, contemplative, and the like. The conclusion was that the human movement tendency is a stylistic variable, and that it does indeed relate, as Rorschach thought it did, to a preference for "intrapsychic living" as opposed to interest in action, practical affairs, and "the world outside."

Received May 20, 1954.

References

1. Arthur, G. *A point scale of performance tests.* New York: Psychological Corp., 1947.
2. Bennett, G. K., & Fry, D. E. *Test of mechanical comprehension.* New York: Psychological Corp., 1947.
3. Gough, H. G. *Predicting success in graduate training.* Berkeley: Univer. of California, Institute of Personality Assessment and Research, 1950.
4. Gough, H. G. *The General Information Survey.* Berkeley: Univer. of California, Institute of Personality Assessment and Research, 1953.
5. Gough, H. G. *A preliminary guide for the use and interpretation of the California Psychological Inventory.* Berkeley: Univer. of California, Institute of Personality Assessment and Research, 1954.
6. Guilford, J. P., Wilson, R. C., Christensen, P. R., & Lewis, D. J. A factor-analytic study of creative thinking: I. Hypotheses and descriptions of tests. *Rep. Psychol. Lab., Univer. Southern Calif.*, 1951, No. 3.
7. Rorschach, H. *Psychodiagnostics.* Bern: Huber, 1942.
8. Terman, L. M., & Oden, M. H. *The gifted child grows up.* Palo Alto: Stanford Univer. Press, 1947.
9. Wesman, A. G. *Personnel classification test.* New York: Psychological Corp., 1947.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Factors of Change in Lobotomized Chronic Schizophrenic Patients

Maurice Lorr
Veterans Administration and Catholic University of America

James Q. Holsopple, Richard L. Jenkins
Veterans Administration

and James P. O'Connor
Catholic University of America

A study to define the nature and extent of changes which occur as a result of the operation of prefrontal lobotomy revealed certain significant changes in lobotomized chronic schizophrenics following a three-month interval (2, 3). These changes were in part measured by scores on 11 factors based on clinical judgments made on an extensive rating schedule (4, 7).

Subsequently, the correlations between the 11 preoperative factor scores of 153 of the patients selected for lobotomy and their matched controls were factor analyzed (5). The analysis was undertaken in the hope of securing a more satisfactory conceptual differentiation of the major schizophrenic processes than is presently available. Three "second-order" factors were identified and related to those reported in previous studies. Factor Alpha was interpreted as a bipolar parameter of apathetic withdrawal with motor disturbances versus an agitated excitement. Beta was regarded as a factor of distortion and disorganization in thinking and perception. Gamma described a fighting reaction which was called belligerence.

The scaled clinical judgments as represented in the 11 factor scores indicated that there was favorable change in both lobotomized and control groups at the end of a three-month period. However, the degree of favorable change was significantly greater for the operated patients on certain of the factor scores. A number of questions arose as to the nature of the changes in the operated patients and their controls. Can change best be described in terms of a single general factor of reduced pathology or in terms of several group factors? If change can be described in terms of several factors, are these the same as those originally identified in the analysis of the preoperative scores? Furthermore, are the change factors the same in the operated and nonoperated groups? If they are similar, then the findings of greater change in the operated patients would suggest that lobotomy simply quickens or stimulates the normal process of recovery. Isolation of different factors in the two groups would imply the development of new patterns or the production of adverse effects or compensatory defects.

In order to throw light on these questions two parallel factor analyses were undertaken, one of changes in lobotomized patients, and one of changes in control patients. The variables used were the changes in 11 factor scores obtained over a three-month interval.

Procedure

The patients were 250 chronic male schizophrenics at six Veterans Administration hospitals. They had been ill for a period of four years or longer. Of the 250 patients, 125 were selected from among a group who were lobotomized and 125 were from a group selected to serve as controls.

The data consisted of clinical judgments recorded on a schedule of 81 brief descriptive rating scales (7). The interview section of the schedule consisted of 46 scales which were completed following a one-hour diagnostic interview by a psychiatrist or psychologist. The ward section consisted of 35 scales filled out by a psychologist upon the basis of information secured by interviewing nurses or attendants who knew the patient well. The schedule itself was designed to record in relatively objective form characteristic behavior and symptoms observable in an interview or on a hospital ward as well as certain items descriptive of hospital adjustment.

Initial ratings were obtained prior to the operation on patients later lobotomized and on patients selected to serve as controls. Three months following the initial description the same patients were again interviewed and rated.

The 11 initial and three-month postoperative scores computed for each patient were based on 11 factors previously identified in a multiple-factor analysis of the rating schedule (6). Each score consisted of the sum of the ratings received by a patient on the group of scales best defining a factor. For example, the paranoid projection factor *C* consisted of the sum of the ratings on scales descriptive of ideas of reference, delusions of persecution, delusions of influence, and frequency of assaults.

The 11 factors scored may be identified by the name and letter that follows:

A Retarded depression versus manic excitement
B Compliance versus resistiveness
C Paranoid projection
D Activity level
E Melancholy agitation
F Perceptual distortion
G Motor disturbances
H Submissiveness versus belligerence
I Aggressive social activity versus social withdrawal
J Self-depreciation versus grandiose expansiveness
K Conceptual disorganization

Each patient's change score, based upon the difference between his initial and his three-month postoperative factor score, constituted the basic data for our analysis. Pearson correlations were computed separately between the 11 factor change scores of the 125 lobotomized patients and the 125 controls.

The 11 × 11 correlation matrix of the operated patients' change scores was factored by the centroid procedure with the highest correlation coefficient in the column utilized as the communality estimate. The communalities obtained were reinserted in the matrix which was again factored. The mean absolute fourth factor residual was then .04 as compared with the standard error of .064 of a zero correlation coefficient.

The correlation matrix of the control patients' change scores was factored with centroid estimates of the communalities in the

Table 1

Correlations Between Eleven Change Scores of 125 Operated and 125 Nonoperated Schizophrenics
(Matrix for operated patients above diagonal; matrix for controls below diagonal.)

Factor scores	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
A		01	08	27	55	06	15	07	−35	17	−03
B	00		38	26	−04	43	41	26	01	37	48
C	17	−03		29	13	50	42	53	−05	25	43
D	20	−16	04		32	33	25	27	−09	23	19
E	34	−09	−02	15		07	14	21	−12	13	07
F	11	06	49	08	−06		27	43	−02	40	51
G	19	26	38	01	07	57		44	36	41	53
H	26	33	21	04	15	13	29		−08	59	31
I	−12	08	08	−33	−03	08	14	−11		02	10
J	21	18	29	04	−03	21	29	44	−03		47
K	−04	13	37	−04	−09	57	53	06	01	17	

Table 2

Centroid Factor Matrices F_o (Operated Group) and F_c (Control Group)

Factor scores	F_o I	F_o II	F_o III	F_o IV	F_o h^2_j	F_c I	F_c II	F_c III	F_c IV	F_c h^2_j
A	40	−67	19	−03	65	42	−37	11	29	41
B	54	29	−09	16	41	19	15	43	−33	35
C	62	18	−25	15	51	49	24	−11	11	32
D	49	−15	−07	17	30	24	−38	−38	10	36
E	39	−54	22	08	50	16	−33	14	36	28
F	62	22	−24	12	50	60	47	−24	16	66
G	60	36	44	14	70	67	40	16	13	65
H	69	18	−12	−34	64	55	−23	37	−26	56
I	−10	43	37	23	38	−11	38	34	25	33
J	62	20	12	−30	53	50	−07	15	−21	32
K	60	39	11	17	55	49	53	−21	−10	58
Sum of squares	3.20	1.47	.59	.41	5.67	2.15	1.33	.78	.57	4.82

diagonals. Four factors were again found to be sufficient. The mean absolute fourth factor residual was .03. The two matrices are shown in Table 1. The rotations were oblique with the object of attaining simple structure.

Results

The centroid factor matrices F_o and F_c for the operated and control groups are shown in Table 2. To facilitate comparisons the rotated factor matrices V_o and V_c are also presented together in Table 4.

Except for factor *I* the first centroid is positive and particularly prominent for the lobotomized group. Implied here is a general factor of "improvement" in the sense of reduced pathology. When schizophrenic patients improve, they tend to improve in all areas. Table 3 presents the mean preoperative and mean postoperative factor scores for the two groups. The changes are nearly all in the direction of less severe symptoms and less deviant behavior. The increase in mean score for the operated patients on factor *A*, retarded depression versus manic excitement, implies movement in the direction of reduced depressive ratings.

Although the mean *A* score of the controls is virtually unchanged, the median score increases from 20.9 to 21.6. However, the factor of aggressive social interest versus social withdrawal (*I*) shows no change for the control group either in central tendency or in the shape of the distribution.

The first factor A_o in Table 2 of the operated group is characterized by motor disturbances (*G*), withdrawal (*I*), and conceptual disorganization (*K*). Reduction in this factor appears to relate to reduction in tendencies toward a catatonic-like withdrawal.

The A_c factor of the control group is defined by withdrawal (*I*), motor disturbances (*G*), and a negative projection from overac-

Table 3

Preoperative and Postoperative Mean Factor Scores of 125 Lobotomized and 125 Control Schizophrenics

Factor scores	Lobotomized group Preoperative mean	Lobotomized group Postoperative mean	Control group Preoperative mean	Control group Postoperative mean
A	21.1	22.6	21.7	21.4
B	7.4	5.9	6.3	6.1
C	12.1	9.5	11.2	10.6
D	12.8	12.4	12.3	12.1
E	20.7	20.0	20.4	20.2
F	12.1	10.0	11.3	10.4
G	18.1	15.4	17.1	16.1
H	11.9	9.8	10.6	10.1
I	18.5	17.1	17.8	17.6
J	14.7	13.5	13.8	12.8
K	21.2	18.9	20.8	19.1

Table 4

Rotated Oblique Factor Matrices
V_o (Operated Group) and V_c (Control Group)

Factor		V_o				V_c			
		A_o	B_o	C_o	D_o	A_c	B_c	C_c	D_c
Manic excitement	*A*	−02	71	−06	04	−03	49	06	09
Resistiveness	*B*	15	−01	41	01	27	−02	−02	55
Paranoid projection	*C*	−02	00	52	−01	01	22	53	04
Overactivity	*D*	01	29	33	−08	−47	05	06	−16
Melancholy agitation	*E*	08	66	01	−03	07	45	−07	−05
Perceptual distortion	*F*	00	−03	50	03	02	22	80	−05
Motor disturbances	*G*	64	24	09	17	28	41	65	26
Belligerence	*H*	−05	−04	14	49	02	20	00	65
Withdrawal	*I*	53	−07	−07	−08	55	26	13	−04
Grandiose expansiveness	*J*	17	06	−01	49	−05	11	16	45
Conceptual disorganization	*K*	37	05	32	08	−01	−02	69	09

tivity (*D*). Represented here, then, is a reduction of withdrawal behavior and an *increase* in motor activity. If A_o and A_c may be regarded as the same factor, this would suggest that recovery from the picture of catatonic withdrawal is more clearly associated with an increase of activity in the unlobotomized patient, while in the patient who has been lobotomized it is more associated with a reduction of conceptual distortion.

The factors B_o and B_c appear to represent schizophrenic excitement. Factor B_o is defined by excitement (*A*) and melancholy agitation (*E*). Factor B_c is defined primarily by excitement, agitation, and motor disturbances. These two factors represent a reduction in schizophrenic excitement. Again, if we assume that B_o and B_c represent the same process, then schizophrenic excitement in control cases is more clearly associated with diminution in motor disturbances.

Factor C_o is characterized primarily by paranoid projection (*C*), perceptual distortion (*F*), resistiveness (*B*), overactivity (*D*), and conceptual disorganization (*K*). C_c is also defined by perceptual distortion (*F*), paranoid projection (*C*), and conceptual disorganization (*K*). However, the projection of conceptual disorganization is substantially larger; and motor disturbances (*G*) is also prominent. Factor C_o may represent a reduction in paranoid projection. Factor C_c seems to involve a more general reduction in disorganization and distortion of perception, thinking, and motor activity than C_o.

Factor D_o appears to represent a grandiose belligerence. In the operated group, this factor is defined by submissiveness versus belligerence (*H*) and self-depreciation versus grandiose expansiveness (*J*). Factor D_c has in addition a reduced resistiveness (*B*). In control patients, therefore, the reduction in belligerence includes a shift in the direction of cooperation.

Table 5 presents the direction cosines between the reference vectors in the two analyses. Both matrices suggest two more general groupings. However, as a minimum of five variables is required to isolate two factors, the correlations between the primaries were not computed or factored. For the operated group paranoid distortion C_o and grandiose belligerence D_o jointly improve and represent a reduction in the fighting reaction and in

Table 5

Direction Cosines Between Reference Vectors
(Cosines for V_o above diagonal; cosines
for V_o below diagonal.)

Factor		*A*	*B*	*C*	*D*
Withdrawal	*A*	1.00	.53	.06	.16
Schizophrenic excitement	*B*	.38	1.00	.25	−.07
Thinking disorganization	*C*	−.24	−.08	1.00	−.18
Grandiose belligerence	*D*	−.01	−.19	−.64	1.00

paranoid reality distortion. In the control group reduction in catatonic-like withdrawal (A_c) and reduction in schizophrenic excitement (B_c) are negatively associated.

It is also of some interest to relate the factors here identified to those found in previous studies. A factor analysis of the correlations between the 11 preoperative factor scores of 153 of the patients included in the present study resulted in the identification of three second-order factors, Alpha, Beta, and Gamma, described earlier in this report (5). Alpha was interpreted as a bipolar factor of apathetic withdrawal with motor disturbances versus excited agitation. Alpha appears to be represented by two change factors. It is the authors' hypothesis that reduction in catatonic-like withdrawal resulted in factors A_o and A_c. Reduction in excited agitation resulted in the change factors B_o and B_c.

Beta of the preoperative analysis (5), X identified by Degan (1), and BB identified by Lorr, Jenkins, and O'Connor (6) are similar to the factors of disorganization and distortion in thinking and perception C_o and C_c. Factors D_o and D_c (grandiose belligerence) of the present study correspond closely to Gamma and identify a fighting reaction.

Summary

The study was undertaken for the purpose of throwing light on the nature of changes resulting from prefrontal lobotomy and routine hospital care on chronic schizophrenic patients. Questions were raised as to whether (*a*) the favorable changes occurring over an interval of three months could be accounted for by one factor or several; (*b*) the changes resulting from lobotomy and hospital care were similar or different; (*c*) the change factors identified were similar to or different from those originally isolated in the same population sample.

The correlations between 11 change scores in 125 lobotomized schizophrenic patients and 125 control schizophrenic patients were independently factored. The change scores were obtained on the basis of preoperative and three-month postoperative ratings on a factored set of 81 rating scales.

Four closely similar factors of change were identified in the two analyses. The schizophrenic processes tentatively identified were reduced social withdrawal with motor disturbances, reduced schizophrenic excitement, reduced grandiose belligerence, and reduced disorganization and distortion of thinking. These findings strongly suggest that although lobotomy results in the improvement of many patients with chronic schizophrenia, the nature of the process of improvement in schizophrenia following lobotomy does not differ greatly from that which may occur without lobotomy. Two of the four factors correspond closely to factors identified in the preoperative study. The other two change factors appear to have emerged out of a third preoperative bipolar factor.

Received June 2, 1954.

References

1. Degan, J. W. Dimensions of functional psychoses. *Psychometric Monogr.*, 1952, No. 6.
2. Jenkins, R. L., & Holsopple, J. Q. Criteria and experimental design for evaluating results of lobotomy. *Res. Publ. Ass. nerv. ment. Dis.*, 1951, **31**, 319–327.
3. Jenkins, R. L., Holsopple, J. Q., & Lorr, M. Effects of prefrontal lobotomy on patients with severe chronic schizophrenia. *Amer. J. Psychiat.*, in press.
4. Lorr, M. Multidimensional scale for rating psychiatric patients. Hospital form. *Vet. Adm. tech. Bull. 10–507*, Nov. 16, 1953.
5. Lorr, M., Jenkins, R. L., & Holsopple, J. Q. Factors descriptive of chronic schizophrenics selected for the operation of prefrontal lobotomy. *J. consult. Psychol.*, 1954, **18**, 293–296.
6. Lorr, M., Jenkins, R. L., & O'Connor, J. P. Factors descriptive of psychopathology and behavior of hospitalized psychotics. *J. abnorm. soc. Psychol.*, 1955, **50**, 78–86.
7. U. S. Veterans Administration. Development of the Northport Record. *Info. Bull. 10–5–21*, Psychiat. Neurol. Issue, July, 1950.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Rorschach Determinant Shift Among Hallucinatory and Delusional Patients[1]

Robert G. Gibby, Bernard A. Stotsky, Robert L. Harrington, and Richard W. Thomas

VA Hospitals, Marion, Indiana and Brockton, Massachusetts

Several studies have shown that the changes which occur from free association to inquiry in the determinants of Rorschach responses can be used as additional sources of information concerning the subject's personality structure and its functioning (2, 3). These changes, which we have called "determinant shift," have value as an additional experimental technique for research with the Rorschach test (1, 2, 3).

In one study (3) psychoneurotics were found to show significantly greater shift from free association to inquiry than psychotics, especially for shading responses. Psychoneurotics showing little shift gave significantly more human movement (M) and form-dominated color (FC) responses and significantly less pure form (F) responses on the free association part of the test than those showing marked shift. Psychotics showing little shift gave significantly more pure color (C) responses; psychotics showing marked shift gave significantly more human movement (M) and color-dominated form (CF) responses.

The increase of shading from free association to inquiry was found to be related to the overt hostility of the examiner (1). Salk[2] found that the extent to which a person elaborated on his free association responses during the inquiry, i.e., the extent of determinant shift, was related significantly with continuance in psychotherapy.

In view of these preliminary findings, it was felt that the determinant shift showed considerable promise as a research technique and deserved further experimental application. This study is another attempt to apply this measure.

Design

Two groups of patients were chosen for this study. The criteria for including a subject in either group were that he be a white, male World War II veteran, have a total IQ of 80 or more on the Wechsler-Bellevue, show no central nervous system involvement, and present definite hallucinatory or delusional symptomatology. The Rorschach was administered to each subject at the time that the hallucinations or delusions were being overtly manifested. In one group were 30 patients who reported definite hallucinations in addition to other symptoms. In the second group were 30 patients for whom there was clear evidence of the presence of a definite delusional system.

We cannot say that in the delusional group there were no hallucinations or that in the hallucinatory group there were no delusions. The absence of material indicating that a hallucinatory patient is delusional or that a delusional patient is hallucinatory does not necessarily mean the absence of these processes. It is, however, possible to say that the characteristic symptom in one group was the presence of a hallucinatory symptom and in the other a delusional symptom.

The persons administering the Rorschachs did not know the nature of the study. All identifying data were removed. The records were scored by one person, whose reliability as a scorer had been established in the previous studies.

---

[1] From VA Hospitals, Brockton, Massachusetts and Marion, Indiana.

[2] Personal communication from Mr. L. Salk.

Results

The two groups were first compared for differences on total Rorschach responses by the conventional method of scoring determinants, which utilizes verbalizations on both free association and inquiry. All comparisons were made for the raw scores and also for the percentage scores for each determinant converted into arc-sine units. This has been described in previous studies (1, 2, 3). The results are summarized in Table 1.

Since there is no difference for the total number of responses, it is possible to compare the groups directly on determinants. Patients with hallucinations give significantly more human movement (*M*) responses, while patients with delusions show significantly more form-dominated color (*FC*) responses and more responses with color as a determinant (Total *C*, including *FC*, *CF*, and pure *C*).

On the free association hallucinated patients give significantly more *M* and *V* responses. No significant differences between the groups are found for the other determinants. Since the frequencies for both *V* and *M* are too small for the differences to be meaningful, the two groups may be considered not to differ significantly on free association.

Table 1

Comparison of Hallucinating and Deluded Patients for Rorschach Determinants

Variable	Means		CR‡	Mean arc-sine differences	CR‡
	Hal.	Del.			
F	14.7	13.6		2.5	
M	1.0	.4	2.6**	6.0	2.9**
FC	.8	1.7	2.3*	−6.1	2.3*
CF	1.1	1.4		−2.5	
C	.5	.4		1.3	
Total *C*	2.3	3.5	1.6†	−7.3	3.5**
FY	.9	1.5		−3.2	
Y+*YF*	.6	.3		2.3	
V	.2	.1		.3	
Total *Y*	1.7	1.9		−1.5	
R	19.7	19.5			

* Significant at the .05 level.
** Significant at the .01 level.
† Significant at the .10 level.
‡ Only critical ratios which are significant at the .10 level or below are reported.

Table 2

Comparison of Hallucinating and Deluded Patients for Rorschach Determinant Shift

Variable	Mean shift		CR‡	Mean arc-sine differences	CR‡
	Hal.	Del.			
F	2.2	3.3	1.6†	−6.4	2.3*
M	.1		6.5**		
FC	.6	1.4	2.5*	−4.0	1.9†
CF	.4	.7		− .9	
C				.9	
Total *C*	1.0	2.0	3.4**	−4.0	1.9†
FY	.7	1.0		−5.7	2.1*
Y+*YF*	.3	.1	1.7†	+ .9	
V	.0	.1	3.3**	−1.1	
Total *Y*	1.0	1.3		−5.9	2.5*

* Significant at the .05 level.
** Significant at the .01 level.
† Significant at the .10 level.
‡ Only critical ratios which are significant at the .10 level or below are reported.

When the two groups are compared on determinant shift, deluded patients show a significantly greater over-all shift in Rorschach determinants. (The shift on pure form is considered the index of total shift.) Deluded patients also shift more on *FC*, Total *C*, and *V*, while hallucinated patients shift more on *M* and *YF*. All differences are significant at the .05 level. For both groups, however, the amount of shift on *M*, *YF*, and *V* is negligible. The differences for these three determinants, while statistically significant, are not psychologically meaningful.

When the arc-sine scores are considered, the deluded patients show significantly more shift for *F*, *FY*, *FC*, Total *Y*, and Total *C*. The differences between groups in shift for *FC* and Total *C* account for the significant differences found between the groups for these same determinants when they are scored by the conventional method utilizing the responses on both free association and inquiry.

Discussion

The differences in determinant shift suggest that deluded patients are more reactive to the stimulus value of the examiner. In the free association situation the two groups show little difference except for *M* and *V*. When by means of the inquiry the subjects are brought into closer interaction with the examiner, the delusional patients respond to a greater extent by elaborating on a greater number of responses. It is noteworthy that the deluded patients show the most shift from pure form to the color determinants. In the past, Rorschach workers have taken the presence of color to indicate capacity for affective rapport and for reacting to external stimulation. The deluded patients appear to show a greater reactivity to external stimulation and more ability to modify their reactions in response to a change in stimulus. They may, however, be experiencing slightly greater anxiety in relating to the examiner (increase in shading). Hallucinated patients give more human movement responses on the free association. This could be interpreted as a tendency on their part to be more concerned with themselves than with external stimuli. Since the frequencies for *M* are small in both groups, this difference may not be particularly meaningful.

In this study the hallucinating patients show more *M* and delusional patients more *FC*. Usually the two determinants are highly correlated and are considered as indicators of emotional stability and ability to control impulses. In this study the correlation of *M* and *FC* is positive but not significant (around .20). Thus, it is not possible from a consideration of *M* and *FC* in this study to determine the relative stability and emotional control of the two groups.

In a previous study (3) it was found that psychotics who shifted markedly appeared to show better emotional control than those who shifted relatively little. If this finding were applied to the present study, it would suggest that the deluded patients show better control in response to the testing situation. Possibly the presence of hallucinations may point to greater impairment of functioning. In support of this possibility one of the authors (4) found that patients who remained hospitalized for more than two years tended to show hallucinations to a greater degree than patients who remitted (chi square significant at .10 level).

The findings reported here are, of course, tentative and subject to further validational investigation. It has been demonstrated, however, that determinant shift is a meaningful measure and that it gives information not always available from the conventional scoring of the total Rorschach response.

The major finding of the study is that the two groups of patients, homogeneous with respect to *R*, differ significantly with respect to determinant shift. Further, when the determinants are scored by the conventional method utilizing both free association and inquiry, the shift from free association to inquiry is responsible for the differences found between the two groups on two of the significant variables, *FC* and Total *C*.

Summary

1. A new method for scoring the Rorschach by means of the shift in determinants from free association to inquiry was utilized, in addition to the conventional method, to study two groups of psychiatric patients. One group consisted of patients whose predominant symptom was hallucinations; in the other the predominant symptom was the presence of a delusional system. There were 30 patients in each group.

2. By the conventional Rorschach scoring method significant differences were found for *M*, *FC*, and Total *C* between the two groups, with hallucinating patients showing more *M* and less *FC* and Total *C*.

3. When compared for determinant shift, the delusional patients showed more shift from free association to inquiry with significant differences for *F*, *FY*, *FC*, Total *Y*, and Total *C*.

4. The findings were related to those of previous studies. It was concluded that patients with delusions tend to react to the stimulus value of the examiner in the testing situation to a greater extent and with better control than hallucinated patients.

Received June 7, 1954.

References

1. Gibby, R. G. Examiner influence on the Rorschach inquiry. *J. consult. Psychol.*, 1952, **16**, 449–455.
2. Gibby, R. G., & Stotsky, B. A. The relation of Rorschach free association to inquiry. *J. consult. Psychol.*, 1953, **17**, 359–364.
3. Gibby, R. G., & Stotsky, B. A. Determinant shift of psychoneurotic and psychotic patients. *J. consult. Psychol.*, 1954, **18**, 267–270.
4. Stotsky, B. A. A comparison of remitting and nonremitting schizophrenics on psychological tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 489–496.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Changes in Wechsler Subtest Scores with Age

Robert J. Howell
Brigham Young University

Since the introduction of the Wechsler-Bellevue Intelligence Scale for adults many investigations have been conducted in an effort to determine the effectiveness of subtest variation as an index of mental deterioration. The purpose of this article is to examine Wechsler's method of establishing normal or average deterioration with age and his assumptions regarding the variability of this deterioration.

Wechsler (6) constructed his tables of normal deterioration by summing the means of the "hold" tests and the means of the "don't hold" tests for each five-year age group (beginning at the age of 20) and dividing the difference of the two by the sum of the means of the hold tests (6, pp. 65, 222, Tables 39 and 40). This method may be thought of as being a *ratio of means*. Undoubtedly, the ease of using data that had already been computed dictated its use. A more conventional manner of determining average deterioration would be to compute deterioration scores for each individual within an age group and then determine the mean of these scores. Such a method may be referred to as a *mean of ratios* score, that is, a mean derived from individual deterioration scores. This latter method permits the computation of variability of deterioration directly, whereas in the former method variability may be estimated only by reference to the variability of individual subtests. Since there is no mathematical relationship between the two methods, any similarity of the resulting means must be determined empirically.

A second question may be raised as to the adoption of a 10% and a 20% deterioration score above that allowed for by normal decline with age as being indicative of possible and definite pathological organic deterioration respectively. Reportedly these figures, 10% and 20%, correspond "roughly to deviations of minus one *PE* and minus two *PE* from the mean at age 20–25" (6, p. 66). How these figures were obtained is difficult to ascertain, since the deterioration score resulting from summing each don't hold test score one *PE* below its mean is 18% at this age interval.

Method

In an attempt to investigate these two problems six subtests of the Wechsler-Bellevue Intelligence Scale, Form I, were administered to 397 American-born white subjects ranging in age from 20 to 89 years. All of the subjects were considered within the normal range so far as behavioral adjustment was concerned. The 20–29 age group was drawn at random from the files of the University of Utah Counseling Center. Being college students, they would be expected to score somewhat above average on the subtests. The rest of the sample was taken primarily from the urban and suburban areas of Salt Lake City. Thirty-eight of the subjects in the 60–89 age group resided in private and charitable homes for the aged, while the remaining 62 lived in the community. All of this group were free from any apparent handicaps that might have affected their performance on the test battery.

The six subtests were Information, Comprehension, and Vocabulary (comprising the hold tests), and Picture Arrangement, Block Design, and Digit Symbol (comprising the don't hold tests). It should be noted that these are not the same tests as suggested by Wechsler for measuring mental deterioration. Originally it was felt desirable to maximize deterioration in studying the nature of the two curves. Inspection of Wechsler's original standardization data (6, p. 222, Tables 39 and

40) shows that the first two tests (Vocabulary is not reported) are the most resistant to increasing age and the last three the least resistant. Other investigators (1, 2, 3, 4, 5) have found the same results. The use of these tests will in each instance yield a higher deterioration score, but since the variability of all the tests is approximately equal at each age level, it should be possible to draw conclusions as to the variability of deterioration as determined by this test battery which would apply to the tests that Wechsler employed in estimating normal deterioration.

Results and Discussion

Table 1 indicates the means and standard deviations of each of the subtests at each age interval. It should be noted that the first three tests show little decline with advancing age, while the last three decline progressively, finally to about half the magnitude of the hold tests. Undoubtedly the older persons' in-

Table 1

Means and Standard Deviations of Subtests

Age	*N*	Info.	Comp.	Vocab.	Pict. Arr.	Block Design	Digit Symbol
20–24	135	12.30 2.00	13.47 1.92	11.88 2.06	12.28 2.06	13.18 2.73	12.61 2.24
25–29	65	13.08 1.88	14.43 1.91	12.58 1.84	12.29 2.94	13.72 2.24	12.51 2.35
30–34	26	12.85 3.21	13.69 2.73	12.77 3.29	10.81 3.25	12.42 3.12	10.92 3.06
35–39	12	11.33 3.40	11.08 4.13	11.08 3.30	9.08 3.50	9.75 3.84	9.75 4.04
40–44	10	11.40 3.67	13.00 3.61	11.40 4.03	8.80 3.19	9.70 3.62	9.10 3.30
45–49	10	12.00 2.32	14.20 2.93	12.80 2.68	9.90 3.14	10.50 3.14	9.10 3.01
50–54	18	11.00 2.16	12.17 2.14	11.50 2.01	8.33 2.11	9.83 2.52	8.17 2.48
55–59	21	10.95 3.33	11.67 3.10	11.28 2.57	7.76 2.65	8.67 2.93	6.90 2.56
60–64	20	11.35 2.52	11.40 2.78	11.80 2.23	7.50 3.01	7.80 2.27	7.05 2.69
65–69	23	11.26 2.75	11.83 2.16	11.52 2.45	6.96 2.31	8.13 2.83	5.87 1.83
70–74	21	10.90 2.89	11.28 1.98	11.38 2.17	5.52 1.62	6.05 1.84	5.43 2.42
75–79	16	10.62 3.16	10.94 2.08	11.38 2.47	5.69 2.44	6.19 2.77	4.50 2.42
80–84	16	9.38 3.08	9.31 2.02	10.81 2.98	4.06 1.34	5.25 2.05	3.31 1.40
85–89	4	9.25 1.05	12.25 2.28	9.00 .71	5.00 1.58	4.50 .50	4.00 1.73

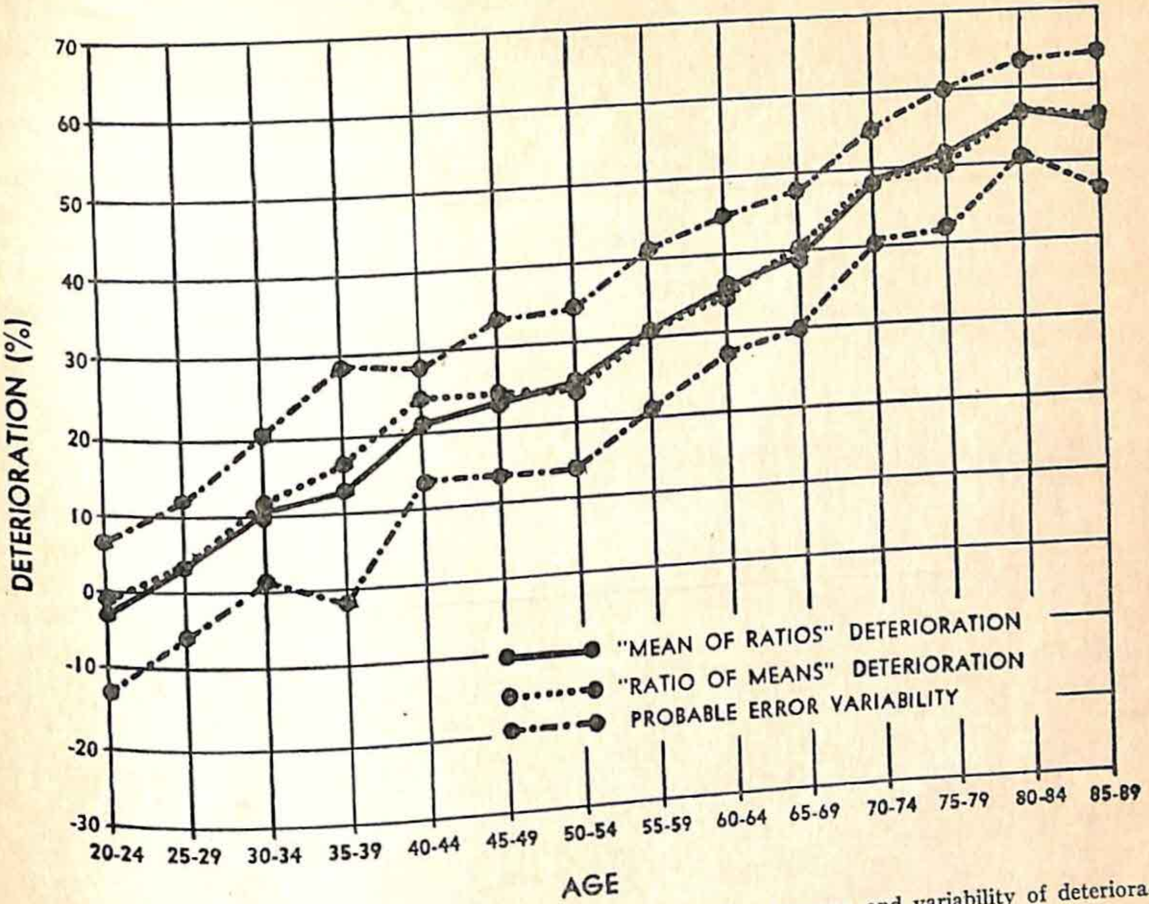

Fig. 1. Percentage deterioration scores by ages: mean of ratios, ratio of means, and variability of deterioration scores.

ability to react rapidly was a contributing factor to the decline, particularly in the case of Digit Symbol; however, time seemed to have little effect on their ability to solve Picture Arrangement and Block Design. Once an individual missed one of the stories or designs, it was unusual for him to correctly complete a subsequent one, regardless of the time involved.

Figure 1 indicates the deterioration curves as computed by each method as well as the variability. It should be noted that the two curves virtually overlap each other at all ages. A plausible explanation of the similarity of the two curves would seem to be in the operation of compensating factors. It should also be noted that the deterioration curves are essentially linear throughout the entire age range. This corroborates Wechsler's findings up to the age of 55 and adds strength to his assumptions of linearity in the older age groups.

It can be seen from Table 2 that the *PE* variability at age 20–24 is 10%, as Wechsler had assumed. Furthermore, inspection of Figure 1 reveals the variability to be approximately constant throughout. In order to test Wechsler's assumption that the *PE* at each age interval was 10% (or 14.8% in standard deviation units) *t* ratios were computed between this theoretical variability and the empirically derived variability (Table 2). In only one instance, age 80–84, was the *t* ratio significant at the .01 level.

A limiting factor in this study was the relatively small number of subjects at each age interval, but it may be tentatively concluded that Wechsler's assumption of equal variability of deterioration at all age levels is correct. It should be noted, however, that no attempt was made to determine the validity of his claim that 10% and 20% deterioration above normal is indicative of

Table 2

Variability of Deterioration Scores at Various Ages

Age	N	Variability (PE)	Variability (SD)	t ratios*	p*
20–24	135	.105	.156	.84	$>.1$
25–29	65	.090	.134	1.17	$>.1$
30–34	26	.101	.150	.10	$>.1$
35–39	12	.159	.235	1.81	$<.1$
40–44	10	.067	.100	2.18	$>.05$
45–49	10	.096	.143	.16	$>.1$
50–54	18	.090	.133	.68	$>.1$
55–59	21	.101	.149	.04	$>.1$
60–64	20	.082	.121	1.42	$>.1$
65–69	23	.089	.132	.84	$>.1$
70–74	21	.075	.111	2.18	$<.05$
75–79	16	.091	.135	.54	$>.1$
80–84	16	.057	.085	4.20	$<.001$
85–89	4	.089	.132	.34	$>.1$

* Significance of the difference between the observed variability and a variability of .10 (10%).

pathological organic damage. This is a clinical problem and not a statistical one.

The assumption of equal variability of deterioration throughout the life span may appear, at first glance, contradictory to the widespread idea that with increasing age there is an increase in variability in both mental and physical dimensions. Such have been the findings of several investigators with respect to intelligence. It may be, however, that the deterioration index is relatively independent of the intellectual level. Likewise the variability of this measure of deterioration may be independent of other seemingly related factors.

Summary and Conclusions

Three hundred and ninety-seven subjects, ranging in age from 20 to 89 years, were administered six subtests of the Wechsler-Bellevue Intelligence Scale, Form I, in an effort to ascertain the similarity of Wechsler's method of computing normal deterioration (a ratio of means) with the more conventional method of determining a mean of individual deterioration scores. In computing deterioration by this latter method it was also possible to determine the variability of deterioration in order to test Wechsler's assumption that the PE of deterioration is 10%, and that this is constant throughout the age span. It was concluded that:

1. Even though there is no mathematical relationship between the two methods of computing deterioration, the resulting curves were virtually identical.
2. Deterioration is linear above the age of 55 as well as in the lower age levels, as found by Wechsler.
3. There is approximately equal variability of deterioration at all age levels.

Received May 13, 1954.

References

1. Chesrow, E. J., Wosika, P. H., & Reinitz, A. H. A psychometric evaluation of aged white males. *Geriatrics*, 1949, **4**, 169–177.
2. Fox, Charlotte, & Birren, J. E. The differential decline of subtest scores of the Wechsler-Bellevue Intelligence Scale in 60–69-year-old individuals. *J. genet. Psychol.*, 1950, **77**, 313–317.
3. Hunt, W. L. The relative rates of decline of Wechsler-Bellevue "hold" and "don't hold" tests. *J. consult. Psychol.*, 1949, **13**, 440–443.
4. Madonick, M. J., & Solomon, M. The Wechsler-Bellevue scale in individuals past sixty. *Geriatrics*, 1947, **2**, 34–40.
5. Rabin, A. L. Psychometric trends in senility and psychoses. *J. gen. Psychol.*, 1945, **32**, 149–162.
6. Wechsler, D. *Measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Davis-Eells ("Culture-Fair") Test Performance of Lower-Class Retarded Children[1]

Sidney Rosenblum, James E. Keller
Wayne County Training School

and Ned Papania
VA Mental Hygiene Clinic, Detroit

It is the contention of Davis and his associates at the University of Chicago (2) that some of the more widely used group and individual tests of intelligence tend to "discriminate" against the lower social class child as a function of the highly verbal, "academic" nature of the items they contain. To remedy this situation, the Davis-Eells Test of General Intelligence or Problem-Solving Ability, a "culture-fair" intelligence test free of reading requirements, has been developed, which, according to its authors, provides a more "valid" estimate of a child's level of intellectual functioning because it draws its problems entirely from experiences common to all or nearly all children who are tested.

The present investigation was designed to evaluate the performance on the Games (as the test is commonly known) of a group of high-grade mentally handicapped (nonorganic) school-aged boys of lower social class standing, and to compare these findings with those obtained from three other tests: a group test of intelligence requiring reading ability (California Test of Mental Maturity), and two individually administered tests of intelligence known to contain verbally loaded, academic-type items (Revised Stanford Binet Test of Intelligence, Form L, and the Wechsler Intelligence Scale for Children).[2] If it is reasonable to assume that many lower social class children diagnosed "mentally deficient" carry such a label mainly because of depressed functioning on intelligence tests supposedly partial to their middle- and upper-class peers, then a group like ours should perform significantly better on a culture-fair test than on the three other instruments under investigation.

Method

Subjects. Thirty boys with a mean chronological age of 12–1 (range 10–2 to 13–8), and a mean initial IQ of 67.57 (range 55 to 75), constituted the experimental population on whom data were collected. Of the group, 15 were Negroes, 15 were white. Mean length of residence in the Wayne County Training School was 2.6 years (range: .5 to 6.7 years).

Procedure. A rating of "low-status" for each subject was determined by Eells' modification of Warner's Index of Status Characteristics.[3] If the Binet, used initially to select the group, had been administered more than six months before, it was readministered. Both the Binet and WISC were administered in counterbalanced order prior to group testing with the Games, which were given approximately three weeks later according to standardized procedure. Five weeks later, the

---

[1] From the Wayne County Training School, Northville, Michigan, Robert H. Haskell, M.D., Medical Superintendent; Thorleif G. Hegge, Ph.D., Director of Research and Education.

[2] Throughout the rest of this study the following abbreviations will be used: WISC–FS, V, P (Wechsler Intelligence Scale for Children—Full Scale, Verbal Scale, Performance Scale); CMM (California Test of Mental Maturity).

[3] In terms of Warner's original categories these subjects would approximate the lower-lower class socioeconomic group. For a detailed description of this modified class typing see Eells and Davis (2, ch. 11).

Games were readministered individually. During this second session an "inquiry" was held on approximately half the test's items (selected in a random manner) to discover if the subjects were choosing right answers for the right reasons.[4] The group California Test of Mental Maturity was administered two weeks later.

Results and Discussion

Table 1 summarizes the comparisons of the mean scores obtained by our population on the Davis-Eells Games and the three other instruments used in this experiment. It is immediately apparent that the mean score on the Games is not significantly higher for our population than those on the other measures under consideration. In one instance, a significantly higher mean score favoring the WISC-P over the Games was obtained. One possible explanation of the WISC-P superiority lies in the actual "kinesthetic" manipulation of things required by most of the WISC-P items. This allows for an appraisal by the subject of his performance, and permits correction of errors as they are recognized. No such self-corrective opportunities are present in answering Games items. The child can respond after one quick glance at the picture and he has no way of "checking" whether his response is right or wrong. It was noted that many of our subjects, although told repeatedly to listen to all the instructions before selecting an answer, were unable to wait until the instructions had been completely read before marking their response. This widespread emphasis on speed and inability to inhibit response tendencies may have contributed significantly to lowered Games performance.

Table 1

Comparisons of Davis-Eells Games with Binet, WISC, and CMM

Test	Mean	*SD*	*t*
Games	64.65	8.02	
Binet	65.87	5.03	.88
WISC			
FS	66.53	6.46	1.57
V	66.63	5.09	1.38
P	72.67	10.09	4.31*
CMM	66.47	7.07	1.24

* Significant at .01 level.

Several explanations can be offered to account for the failure of the Games to reveal the "hidden" potential in our population supposedly not tapped by other tests of intelligence. First, though the test may not contain culturally biased items, as Davis and Eells contend, the conditions under which the test was administered to our group cannot be said to have been completely free of middle-class cultural overtones. The Games were given to our subjects in a school building on a school day by middle-class individuals undoubtedly identified with the school system. Although the group was solving problems via pictures rather than responding to a series of academic-type items, the materials used in this process were pencil and paper, the basic physical ingredients involved in most school work. Although presented to them as a series of games, which Davis and Eells claim will encourage a child to exert maximum effort, our subjects were not fooled and revealed by their comments that they knew from the outset that they were taking another test. Since it is known that an academic atmosphere often elicits more feelings of hostility and lowered interest on the part of many lower-class children as compared to their more scholastically oriented middle-class peers, it might be reasonable to assume that similar feelings were aroused in our subjects during administration of the Games, and that these feelings had a negative effect on their performance. In other words, what is being suggested is that the method of group administration of the Games did not elicit the highly motivated behavior on the part of our subjects that might be expected from a sample of their middle-class peers.

An alternative explanation for our findings is that our subjects are genuinely deficient in problem-solving ability, and that this deficit would manifest itself not only in a middle-class culture but in their own lower-class cultural milieu as well. Analyses of the data obtained from individual inquiries lend support to such a hypothesis. Our subjects made errors on the Games not because they were unfamiliar with the content of the items but because of a failure to make correct associations, discriminate crucial differences, and organize elements of a problem into a meaningful pattern.

---

[4] No analyses of the data obtained from the individually administered Games are presented since the writers' sole purpose was to investigate the "validity" of the subjects' responses through an inquiry, which could be done only through individual sessions with each child.

Table 2

Percentage of Subjects Giving Correct Reasons for Correct Responses on the Davis-Eells Games

Item	Percentage
Probabilities	10
Money	0
Best Ways	47
Analogies	3
All items	0

Table 2 gives a tabulation of the percentage of subjects giving correct reasons for their correct Games responses. It can be seen that not one of our 30 subjects was able, upon questioning, to verbalize the reasoning he should have used for all cases for which he chose the correct answer. Davis and Eells cite a study (4) in which it is indicated that 92% of a group of normal school children who marked right answers for the "analogy" problems explained the analogous relationship correctly in an interview. The figures of Table 2 indicate that only 3% of our population were able to do this, and that not a single subject used the correct approach in solving the "money" problems. Both these tests demand an abstract approach and require the subject to use a series of intervening symbolic mental steps in solving the problem. "Best ways" and "probabilities," which seem to require a less abstract set, were handled somewhat better, but even with these tests the percentage of subjects giving correct reasons for correct answers is not very high.

In general, then, these data tend to support the now widely accepted view that one very important condition in cases of mental deficiency is the lack of ability to abstract. It may be, as McCandless (3) and others have indicated, that membership in the lower social classes tends to produce this concretistic mode of thinking, and that our group represents an extreme example of this phenomenon. Whatever the explanation, our study suggests that performance on the Davis-Eells Games is as much affected by this loss in abstract thinking ability as are other tests of intellectual functioning.

In closing, it should be pointed out that the experimenters have been concerned only with the ability of the Davis-Eells Games to tap an intellectual potential not revealed by other standardized tests of intelligence. The present investigation strongly suggests that they do not do so with our population. However, the test may prove useful in another way—by predicting the ability of children to handle the practical problems peculiar to daily Training School living that are not revealed or suggested by school achievement or routine test performance. The authors of the Games state their test requires the subject to use mental activities also required in "solving most problems in life and in the school curriculum" (1), but to date no work has been published to demonstrate the validity of their assumption either with an institutionalized or noninstitutionalized population. It is one thing to prepare an operationally sound test to *define* a concept such as problem-solving ability; it is entirely another matter whether the test is *useful* in the sense that it predicts the behavior it purports to measure. As yet, we do not know the types of behavior the Games predict for our specific population. Further research will be needed to answer this question and to investigate the potentialities of this test for use with the high-grade mentally handicapped child.

Summary

To determine if the Davis-Eells (culture-fair) Games would tap an intellectual potential not revealed by other tests of intelligence and reveal a "truer" picture of the level of functioning of a group of lower-lower social class, high-grade mentally handicapped (non-organic) children, the Games were administered to 30 school-aged boys in residence at the Wayne County Training School. Results

were compared with those of three other standardized intelligence tests: the Binet, WISC, and California Test of Mental Maturity.

In general, it was concluded that for this population the Davis-Eells Games do not tend to reveal a "hidden intellectual potential"—by virtue of their elimination of culturally unfair items—not tapped by other intelligence tests presumed to be culturally biased. Several explanations are offered for this finding, and suggestions for future research with the Games are indicated.

Received June 10, 1954.

References

1. Davis, A., & Eells, K. *Manual for the Davis-Eells Test of General Intelligence or Problem-Solving Ability*. Yonkers, N. Y.: World Book Company, 1953.
2. Eells, K., Davis, A., Havighurst, R. J., Herrick, V. E., & Tyler, R. W. *Intelligence and cultural differences: a study of cultural learning and problem-solving*. Chicago: Univer. of Chicago Press, 1951.
3. McCandless, B. R. Environment and intelligence. *Amer. J. ment. Def.*, 1952, **56**, 674-691.
4. Weiss, Marcia. The validity of nonverbal problems from a culture-fair test for middle and lower-class children. Unpublished master's thesis, Univer. of Chicago, 1952.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

A Sexual Deviation Scale for the MMPI

James T. Marsh, Jessamine Hilliard, and Robert Liechti
School of Medicine, University of California at Los Angeles

This report represents the preliminary findings in the development of a sexual deviation scale for the MMPI. The work was done as one phase of a research into biochemical and personality factors in sexual deviation.[1]

This tentative MMPI scale was constructed in answer to the question whether or not individuals convicted of sex offenses can be distinguished on the basis of an objective personality test from presumably normal individuals having no records of arrests for sexual offenses.

A scale capable of this discrimination could be a useful adjunct to established clinical procedures in the diagnosis and treatment of sexual psychopathy. In addition, such a scale could find application as a screening device in the selection of individuals for work involving close contact with children.

The MMPI was chosen as the instrument for study since it is objective, easily administered and scored, and includes a wide range of items tapping attitudes, affects, personality traits, and symptoms. The variety of items included in the MMPI in particular represents an advantage for a scale of this type, since it makes possible discrimination on the basis of items not directly related to sexual attitudes or experiences, and thus may be less accessible to modification by the defensive subject who seeks to present himself in the most favorable light.

Subjects

Sex offenders. The sex offender population was composed of men convicted of some sex offense and committed to the Metropolitan State Hospital for observation, diagnosis, and treatment under the provisions of the California Sexual Psychopath Act. They ranged in age from 17 to 70 years, with a mean age of 36 years. Eighty-five per cent were native-born white men, and approximately 11 per cent were of Spanish-American origin. Intelligence levels, as measured by the Shipley-Hartford or Beta tests, were approximately normal in distribution although slightly skewed toward the higher intelligence ranges. Almost 60 per cent of the group had records of one or more previous convictions, two-thirds of which were for sexual offenses. Acts leading to hospitalization under the Sexual Psychopath Act include indecent exposure, child molestation, rape, and homosexual contacts with minors, the majority of these offenses involving sexual activity with children.

All sex offenders committed to the hospital for observation and diagnosis are given a battery of psychological tests which includes the group form of the MMPI. The tests are administered within the first few weeks after admission. The MMPI records utilized in the present study were obtained from the hospital files.

Comparison subjects. These subjects were drawn from a population of male students in education at U.C.L.A. They ranged in age from 18 to 47 years, with a mean of 30 years. The group was composed both of full-time undergraduate university students and individuals active in the field of education who were doing graduate work toward advanced degrees. The latter included superintendents, principals, and teachers from the Southern California area. The mean intelligence and education levels for this group exceeded that of the normal population. These students are routinely given a battery of psychological tests including the MMPI during the course of training. Records used in this study were drawn from the files of the education department at U.C.L.A.

---

[1] Conducted jointly by the Medical School of the University of California at Los Angeles and the Metropolitan State Hospital, supported by grants from the legislature of the State of California.

Procedure

Construction of the scale was begun by drawing 100 MMPI records in the form of machine-scoring answer sheets from the hospitalized offender population and 100 records from the normal population. These were subjected to item analysis. A complete count was made of the true and false responses on each of the 566 MMPI items for both normal and sex offender groups. The number of true and false responses for each group on each item was then tabulated and this tabulation carefully inspected item by item for differences between the two groups. In each instance where considerable difference in the proportions of true and false answers appeared, a chi-square test was applied to determine the probability that such differences may have occurred by chance. In those instances where the number in any cell of the fourfold chi-square table fell below five, Yates' correction for continuity was applied. As a result of statistical analysis, 100 items were selected, each of which differentiated the normal and sex offender groups at a statistically significant level of .01 or beyond. In order to conserve space, only the numbers of these items, taken from the MMPI group form booklet, are reproduced here.[2]

Sex offenders, in contrast to the comparison group, tend to answer as true items 5, 59, 61, 67, 76, 84, 88, 94, 98, 106, 111, 118, 127, 138, 139, 140, 144, 147, 158, 168, 170, 179, 239, 249, 260, 297, 298, 303, 304, 316, 329, 346, 348, 349, 350, 352, 360, 364, 365, 373, 375, 377, 378, 382, 385, 389, 395, 404, 413, 419, 427, 444, 453, 455, 457, 458, 483, 488, 489, 490, 492, 498, 507, 548, 549, 558, 559, and 562.

Sex offenders, in contrast to the comparison group, tend to answer as false items 6, 12, 20, 37, 39, 46, 51, 63, 89, 120, 133, 134, 155, 160, 163, 255, 289, 294, 302, 328, 347, 367, 372, 376, 379, 380, 430, 432, 449, 460, 547, 554.

To test the possibility that these 100 items might be of discriminative value, an MMPI scoring key was constructed from the responses of the sexual offender group. Scores for these items were then obtained for 200 additional sex offender records and 145 new student records. Each response in a deviant direction (i.e., answered in the same direction as the original sex offender group answered it) was scored as one unit with the possible range of scores extending from 0 to 100. The actual range of items answered in a deviant direction for this total sample of 345 records fell between 10 and 70.

Results

The distribution of scores for the sex offender and comparison groups obtained on the new scale is presented in Table 1. The mean number of items answered in the deviant direction by the hospitalized sex offenders is 42.1 with a standard deviation of 9.3, while that of the comparison student group is 21.3 with a sigma of 6.5. The best cutoff point to distinguish these groups seems to lie between 30 and 31 items.

For cross-validation purposes, two new samples were drawn from the sex offender and comparison populations and scored on the sexual deviation key. These samples con-

Table 1

Distribution of 145 Normals and 200 Sex Deviates on a 100-Item Sexual Deviation Scale

MMPI scale score	Frequency	
	Normals ($N=145$)	Sex offenders ($N=200$)
66–70		
61–65		2
56–60		3
51–55		14
46–50		26
41–45		31
36–40	2	30
31–35	4	36
26–30	12	38
21–25	22	9
16–20	38	9
11–15	39	2
6–10	25	
0–5	3	
Mean	21.3	42.1
SD	6.5	9.3

---

[2] An 8-page table giving the number, text, deviant answer, χ^2, and p level of each item has been deposited with the American Documentation Institute. Order document number 4420 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

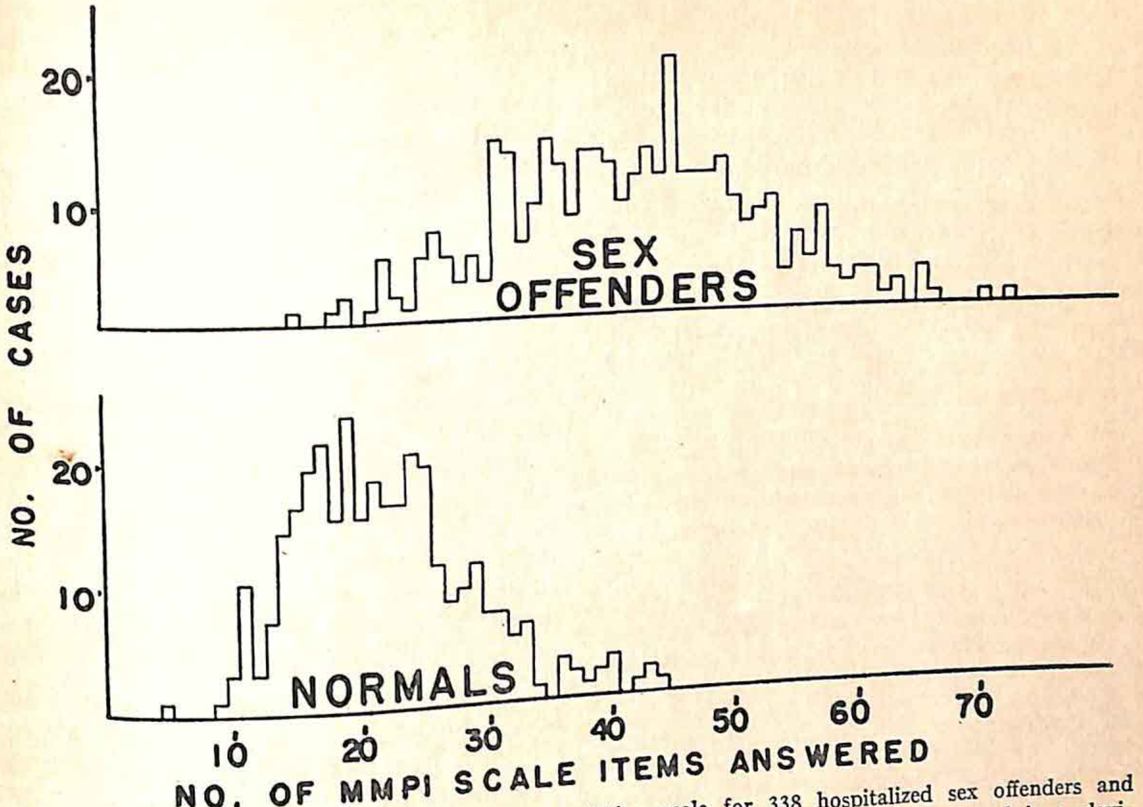

Fig. 1. Distribution of scores on MMPI sexual deviation scale for 338 hospitalized sex offenders and 317 students. The abscissa represents the number of items from the 100-item scale answered in a deviant direction.

tained 138 sex offender records and 172 student records. Distribution of cases and means in these new groups so closely approximated that of the original samples that it was possible to combine validation and cross-validation data.

Figure 1 illustrates the graphic distribution of all combined samples, 317 normals and 338 offenders. The mean number of items answered in a deviant direction by members of the comparison group is 21.8 with a standard deviation of 6.8, while the sex offender mean is 41.7 with a standard deviation of 9.1.

Table 2 presents the distribution of the combined validation and cross-validation samples on the new scale. Scores of 283 normals and 41 sex offenders fall below a cutoff point placed between 30 and 31 items, whereas scores of 34 normals and 297 sex offenders fall above. This scale appears to differentiate these groups with a high degree of efficiency since it correctly identifies 88 per cent of offenders and misclassifies only about 11 per cent of the comparison subjects.

Table 2

Distribution of 317 Normals and 338 Sex Offenders on the Sexual Deviation Scale

MMPI scale score	Frequency	
	Normals ($N=317$)	Sex offenders ($N=338$)
71–75		1
66–70		2
61–65		9
56–60		20
51–55		33
46–50		54
41–45	4	64
36–40	11	58
31–35	19	56
26–30	46	23
21–25	89	14
16–20	93	3
11–15	50	1
6–10	4	
1–5	1	
Mean	21.8	41.8
SD	6.8	9.1

Discussion

With a view to gaining a clearer understanding of factors contributing to high and low scale scores, cases falling at the extremes of the two distributions and in the area of overlap were carefully inspected. Screening interview notes and psychological test data were available for the education student comparison subjects with high and low scale scores, and upon inspection of these it was evident that personality difficulties had been noted in almost every case where sexual deviation scale scores exceeded 35 items. The following notes are typical of many cases in this range:

> Scale score 39. "Referred for psychiatric counseling with history of extreme irresponsibility. Lives in parasitic relationship with mother figure. In interview stated inability to form normal relations with either sex. Abandoned by mother very early. Stated he wished to talk to a male rather than a female doctor. MMPI: *K* 78, *D* 84, *Pd* 86, *Mf* 98, *Pt* 95, *Sc* 96, *Ma* 86. A very serious teaching risk who will probably be barred."
>
> Scale score 34. "Referred for psychiatric counseling. Psychosomatic skin disorder. Wishes further therapy and now in treatment. MMPI scores above 70: *D* 75, *Mf* 78, *Pt* 76, *Sc* 78."
>
> Scale score 36. "Referred for psychiatric counseling. Math major, chess enthusiast. Lives with aunt and mother. Bright but socially inept. Difficult family situation resulted in depression and 'nervous breakdown.' Approved for teaching after further testing. No MMPI scores above 70."
>
> Scale score 44. "Referred for psychiatric counseling. Married, good skills. Very religious, recently converted. Helped much by wife after 'nervous breakdown' following two navy enlistments. MMPI scores above 70: *Mf* 85, *Hs* 75."

The case history material from the student group suggests that scores of 35 and over may be indicative of personal maladjustment. With the information available, it was not possible to ascertain either the presence or absence of deviant sexual behavior and the relation of this factor to high scores in the student group is unknown. Inspection of case records of students scoring in the low ranges (10 to 20) revealed a considerably lower incidence of personality maladjustment.

Sex offenders with high and low scale scores. Psychiatric histories and probation reports on the sexual offenders whose scores fell at the extremes of the scale were also inspected. Child-molesting offenses were found to occur more frequently in the group with highest scale scores. A higher incidence of cases diagnosed as cerebral arteriosclerosis was also found in this group. In addition, these individuals had completed on the average fewer years of schooling, and tended to be older than those with the lowest scores. When MMPI scale scores were compared for these extremes, it was found that those with high sexual deviation scores also had significantly higher scores on the *F*, *D*, *Pa*, *Pt*, and *Sc* scales. A significant difference was also found for the *K* scale, high deviation scores being associated with low *K* values.

Scale scores of neurotic and psychotic patients. As a further check on the discriminative value of the scale, MMPI records of 50 neurotic and 50 psychotic patients at Metropolitan State Hospital were scored with the sexual deviation key. Scores ranged from 21 to 60 items answered in a deviant direction. In general, psychotic patients received higher deviation scores than neurotics, their mean falling at 39, whereas the mean for the neurotics was 36. The distribution of scores suggests that the scale does not discriminate sex offenders from hospitalized mental patients. The overlap in the distribution of scale scores for these two groups lends support to the hypothesis that some factor of personality integration or adjustment is being measured by the scale. Scores from both groups appear to be distributed in accordance with such a factor since the lowest scores in the psychotic group were obtained from patients diagnosed as paranoid schizophrenics whereas the highest scores were from catatonics and hebephrenics. This is compatible with the clinical finding that paranoid schizophrenics as a group retain a better reality contact than other schizophrenics.

MMPI items composing the sexual deviation scale. The majority of the items which differentiate hospitalized sex offenders from comparison subjects can be grouped into categories in accordance with attitude or feeling content. The pattern of these items, although not descriptive of any single individual, yields a rough, self-descriptive picture of the hospitalized sex offender which agrees quite closely with clinical observations reported in the literature.

> Sex offenders are often characterized as fearful, passive-dependent, inadequate individuals who are sensitive, shy, and inept in social relations. MMPI items 39, 111, 138, 140, 147, 304, 389, 395, and 459

are generally answered by sex offenders in the direction which illustrates this point. Poor social relations are particularly evident by their answers to items 59, 348, 377, 404, 449, 455, and 547. A tendency to be suspicious and to project unacceptable thoughts and feelings on others is shown by their answers to items 127, 347, 364, 376, and 507. That personality difficulties may date back to relationships and experiences in childhood is suggested by items 118, 419, 453, 458, and 562. As a group, hospitalized sex offenders are subject to many fears which are often unrealistic, as illustrated by items 5, 352, 360, 367, 385, 492, and 559. They suffer from a variety of mental and physical complaints as indicated in items 51, 63, 139, 155, 160, 163, 168, 346, 349, 350, and 365. Guilt and depression, probably related to deviant sexual behavior, are prominent feelings, as is evident in items 49, 61, 67, 76, 84, 106, 158, 375, 379, 382, 413, and 489. The guilt and depressive reactions may stem in part from violation of taboos laid down by strong fundamentalist religious convictions reflected in items 98, 249, 373, 483, 488, and 490. The focus of conflict in the sexual area is evident in items 20, 37, 133, 179, 297, 427, 548, and 558. Some asocial and amoral attitudes are apparent in items 144, 289, 294, 298, and 316.

The variety of categories into which items can be grouped points to a broad base for the discriminative power of the scale.

Limitations of the scale. Despite the promising preliminary results obtained with the scale, its usefulness for diagnostic purposes is restricted at present by the selective nature of the validation groups. The hospitalized sex offender group cannot be considered as representative of the total population of sex offenders since several factors operate to screen this group. One screening factor is the ratio of offenses to arrests. It has been estimated that only a fraction of sex offenses results in arrests. Further, the majority of individuals arrested and convicted of sex offenses are not committed to hospitals for care and treatment under the California Sexual Psychopath Act. Few individuals are committed on the basis of first offenses unless these involve aggressive behavior or minor victims. In view of these selective factors the sex offender group used in this study may not be representative of the total population of individuals who behave sexually in ways which violate the law. It does, however, represent a group of individuals who, in the best judgment of psychiatrists and the courts, must be considered dangerous to the health and safety of others because of aberrant sexual behavior.

The comparison group drawn from a student population is also subject to selective factors. Certainly, this group is atypical of the general population in terms of educational level and possibly atypical in terms of socioeconomic background. The fact of college attendance presumes also a degree of sophistication in terms of tests which the majority of sex offenders have not had the opportunity to acquire.

In view of these limitations, diagnostic application of the scale should await further validation with new groups. In particular comparison groups with socioeconomic, educational, and experiential backgrounds similar to the sex offender group are needed. The scale is presented at this time in the hope that other interested investigators may extend its range of application through further validation.

Summary

In the effort to develop an MMPI scale to distinguish sex offenders from normals, responses of 100 hospitalized sex offenders and 100 normal education students to 566 MMPI items were subjected to item analysis. One hundred items were found which differentiated the two groups at a statistically significant level in terms of direction (true or false) of response.

These items were combined into a scale for which scoring keys were constructed. The scale was applied to new samples of 338 hospitalized sex offenders and 317 education students. Utilizing a cutoff point of 30 out of a possible 100 items answered in a deviant direction, the scale was found to discriminate between groups with but 12 per cent false negatives among sex offenders and 11 per cent false positives among normals.

Cases with scores at the extremes of the distributions of both groups were examined for possible factors contributing to the scores. The items discriminating offenders from students were combined into clusters which describe some personality features of sex offenders.

Received May 10, 1954.

References

1. Guttmacher, M. S. *Sex offenses, the problem, causes and prevention.* New York: Norton, 1951.
2. Hathaway, S. R., & McKinley, J. C. A Multiphasic Personality Schedule (Minnesota): I. Construction of the schedule. *J. Psychol.,* 1940, **10**, 249–254.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

The Relationship of the Wechsler-Bellevue Verbal Scale and the Progressive Matrices Test[1]

Mahesh M. Desai
Belmont Hospital, Sutton, Surrey, England

In view of the fact that the Progressive Matrices (PM) and the Wechsler-Bellevue (WB) are both regarded as good measures of intelligence, the two tests are compared in this study. A group of 190 consecutive male admissions to the hospital were given the PM as a group test during the second week after admission, and within the following 3 days they were given 3 or more tests of the WB Verbal Scale (Form I). Of these 120 had all 6 tests. The mean age was 31.7 (*SD*, 8.8) and the socioeconomic status almost wholly lower-middle or working class. Psychiatrically, 121 were neurotics, 30 psychopaths, 13 organics, 9 early schizophrenics, and 17 miscellaneous.

Product moment *r*'s were computed between PM and WB weighted scores for the full scale as well as for each test and changes in the grades attained were also computed. The mean score on the PM was 43.3 (*SD*, 8.9) and for WB Verbal Scale 51.96 (*SD*, 11.8). Information (*I*), Comprehension (*C*), Digits (*D*), Arithmetic (*A*), Similarities (*S*), and Vocabulary (*V*) had mean scores of 10.8, 11.2, 9.4, 10.6, 10.3, and 11.4, respectively. The *r* of PM with WB was .573, which rose to .648 when corrected for attenuation. The *r*'s of the subtests with PM were *I*, .33; *C*, .47; *D*, .41; *A*, .49; *S*, .50; and *V*, .34, which rose to *I*, .37; *C*, .57; *D*, .52; *A*, .65; *S*, .62; and *V*, .38, when corrected for attenuation. A change of grades based on percentiles was found in 109 *S*s, 88 of whom were higher and 21 lower by one grade or more in WB than in PM.

The over-all *r* is not as high as might be expected in two very good measures of *g*. Even after correction for attenuation the *k* is still .762, making E no more than 23.8. *V* and *I* have the lowest *r*'s in our sample. Part of the explanation for this might be in the relatively lower reliability of PM in psychiatric cases and in the fact that *V* and *I* are less vulnerable to age or psychiatric illness than other WB tests. The last fact might also in part explain the improvement in 88 *S*s. The differences in the structures of the two tests are probably very important in explaining the relatively low *r* and the change in gradings, PM being a measure of present efficiency and involving eduction of relations and correlates, WB largely measuring *g* through what it has already achieved in the past. The highest correlations of *A* (Arithmetical reasoning) and *S* (eduction of relations) with PM might also be partly explained on this basis. Other variables which may have influenced the disparities are the effects of the sequence in which the tests were given, the mode of administration (group vs. individual), and differences in the heterogeneity of the tests which may have affected attitudes, interest, and concentration.

Brief Report
Received November 9, 1954.

---

[1] An extended report of this study may be obtained without charge from Mahesh Desai, Belmont Hospital, Sutton, Surrey, England, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4436 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

The Relationships Among Five MMPI "Anxiety" Indices

Charles Windle
Human Research Unit No. 3, Fort Benning, Georgia

Within the last twenty years anxiety has become widely recognized as a central concept in psychology (9). Introduced and popularized by psychoanalysts (3, 8), this concept has been adopted even by experimental psychologists to describe motivation (10, 13, 17).

Unfortunately, the widespread acceptance of the construct of anxiety is not paralleled by unanimity in definition. The disagreement and confusion that exist in defining this concept are exemplified by the variety of anxiety indices proposed for the MMPI. Five such MMPI anxiety scores have come to the author's attention.

1. Modlin observed that patients classified as having anxiety neuroses were similar in having *Hs*, *D*, and *Hy* as their highest MMPI scale scores. He suggested that the combined score of these three categories be called "the anxiety score (*A* score)" (11, p. 762). Shoben, also, has used these three MMPI scales, the neurotic triad, as a measure of anxiety, arguing that "anxiety is the central factor in most cases of maladjustment and neurosis" (16, p. 42).

2. Purcell *et al.* questioned whether the *Hy* scale is relevant to the detection of anxiety. They replaced the *Hy* with the *Pt* scale as one more likely to measure "the vague fears, over-concern, and worry characteristic of an anxiety reaction" (14, p. 84). Their method of scoring, in which all scores less than 60 are given zero weight, is obviously unsuitable for "normal" populations. The simple sum of the *Hs*, *D*, and *Pt* scales promises to be a more generally useful score and will be employed in the present study.

3. Welsh felt that Modlin's index overlooked two basic features of the anxiety profile, "the anticlinal relation of *D* to *Hs* and *Hy*, and the secondary rise on *Pt*" (22, p. 66). His proposed formula,

$$\text{Anxiety Index} = \frac{Hs + D + Hy}{3} + [(D + Pt) - (Hs + Hy)]$$

was found to follow clinical expectations when applied to various psychiatric and nonpsychiatric groups.

4. Winne proposed a 30-item scale of neuroticism based on a contrast between the responses of neurotic and normal veterans. Only items from the *Hs*, *D*, and *Hy* scales were used. In describing this scale Winne noted "the diffuse or free floating anxiety measured by this adaptation of the MMPI" (24, p. 120). It seems reasonable to assume that Winne conceived of this scale as a measure of anxiety. Holtzman *et al.* (7) have made this same interpretation and even used Winne's scale to "validate" the Taylor anxiety scale.

5. Taylor (19, 20) constructed a "manifest anxiety scale" from items which clinicians agreed were indicative of manifest anxiety.[1] Her scale was developed as a device for selecting experimental subjects hypothesized to differ in "state of *reactivity* or *excitability*, which in turn reflects their general level of drive (*D*)" (2, p. 120). Numerous studies (2, 12, 15, 17, 18, 19, 21) have presented evidence that extreme groups selected on the basis of Taylor anxiety scale scores differ in performance in a manner predictable from Hullian theory.

To date, there is little evidence for preferring any of these anxiety indices (with the possible exception of the Taylor scale) over any others, nor is there adequate evidence as to whether relationships found with one index would also be found with others. Previous studies (1, 4, 7) have been limited to inter-

---

[1] As Holtzman *et al.* (7) have observed, the number of items used in computing Taylor anxiety scores has varied considerably in different experiments. The same appears to be true for the particular items used. The scores in the present study were based on the list of 50 items presented by Taylor (20, p. 286). The MMPI numbers of the items used are: 7, 13, 14, 18, 23, 31, 32, 43, 67, 86, 107, 125, 142, 158, 163, 186, 190, 191, 217, 230, 238, 241, 242, 263, 264, 287, 301, 317, 321, 322, 335, 337, 340, 352, 361, 371, 397, 407, 418, 424, 431, 439, 442, 499, 506, 523, 528, 530, 549, 555.

Table 1

Intercorrelations and Reliability of Anxiety Scores of 55 Females on Two Administrations of the MMPI

Anxiety measures	Purcell's	Welsh's	Winne's	Taylor's	Mean correlation	Retest reliability coefficient
Modlin's	.68(.52)*	.36(.08)	.35(.27)	.37(.31)	.45(.30)	.80
Purcell's		.86(.79)	.54(.63)	.83(.82)	.75(.71)	.91
Welsh's			.33(.44)	.74(.61)	.63(.53)	.84
Winne's				.56(.65)	.45(.51)	.79
Taylor's					.66(.63)	.92

* The first coefficients are from the first test administration; the coefficients in parentheses are from the second administration.

relating only three of the five MMPI anxiety indices. In addition to their limited scope, these studies suffer from minor inaccuracies in the indices used, and from the need for cross validation.

The present study proposed to determine the interrelationships of all five anxiety indices within a single population, incidentally providing an estimate of retest reliability for each index.

Method

Fifty-five female undergraduate students at the State University of Iowa were given the MMPI group form twice, with a one-week interval between tests. The subjects were given no explanation of why they were being either tested or retested, other than that these tests were part of an experiment, nor were they given their test scores. Correlations among the five anxiety indices proposed for the MMPI were computed.

Results

The results of intercorrelating the five anxiety scales are presented in Table 1. In general, Purcell's score was the most highly related to the other anxiety indices. The highest single relationships appear to be between the Purcell index and both Welsh's anxiety index and Taylor's anxiety scale. It is noteworthy that these three scores also have the highest retest reliability coefficients. In contrast, Modlin's *A* score and Winne's neuroticism score showed generally low relationships with other scales, and had relatively low retest reliability coefficients. Item overlap does not seem sufficient to account for the results of the study. The greatest overlap among these five scales is between Modlin's *A* score and Purcell's index, whereas the correlation between these scales is of only intermediate value.

Discussion

Previous investigators have presented evidence both for (4, 5, 7) and against (1, 4, 5, 23) high relationships among anxiety measures. The present study also yields evidence both for and against high relationships. The high correlation of .74 between the Taylor anxiety scale and Winne's neuroticism score claimed by Holtzman *et al.* (7) was not completely confirmed in the present study. Neither Winne's neuroticism score nor Modlin's *A* score was sufficiently highly correlated with the other scales to assume equivalence among them. This conclusion appears doubly warranted in view of the item overlap between indices and the elimination of changes within subjects over time, a usual source of error variance in correlational studies.

On the other hand, higher correlations than have previously been claimed were found between Purcell's anxiety index and both Taylor's anxiety scale and Welsh's anxiety index. The magnitude of these correlations, ranging from .79 to .86, is high enough for the user to assume the equivalence of the scales involved.

Purcell's score was the most highly related to each of the other anxiety indices. It would appear that if there is any unitary anxiety factor measured by all of these scales, Purcell's index best measures that factor. Unfortunately, a factor analysis to determine more accurately the relationships among these five anxiety measures is inappropriate because of the considerable overlap of items among the different scales (6).

Summary

The intercorrelations among five "anxiety" measures from the MMPI were determined for both test and retest of 55 undergraduate females. Three of the MMPI anxiety measures seem fairly highly interrelated, while two had relatively low intercorrelations. In all, Purcell's index, the sum of the *Hs, D,* and *Pt* scales, tended to have the largest correlations with other anxiety indices.

Received May 19, 1954.

References

1. Bitterman, M. E., & Holtzman, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
2. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
3. Freud, S. *The problem of anxiety*. New York: Norton, 1936.
4. Gallagher, J. J. Manifest anxiety changes concomitant with client-centered therapy. *J. consult. Psychol.*, 1953, **17**, 443–446.
5. Gleser, Goldine, & Ulett, G. The Saslow Screening Test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8**, 279–283.
6. Guilford, J. P. When not to factor analyze. *Psychol. Bull.*, 1952, **49**, 26–27.
7. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853–854.
8. Horney, Karen. *New ways in psychoanalysis*. New York: Norton, 1939.
9. May, R. *The meaning of anxiety*. New York: Ronald, 1950.
10. Miller, N. E., & Dollard, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
11. Modlin, H. C. A study of the Minnesota Multiphasic Personality Inventory in clinical practice. *Amer. J. Psychiat.*, 1947, **103**, 758–769.
12. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
13. Mowrer, O. H. *Learning theory and personality dynamics. Selected papers*. New York: Ronald, 1950.
14. Purcell, Claire K., Drevdahl, J., & Purcell, K. The relationship between altitude-I.Q. discrepancy and anxiety. *J. clin. Psychol.*, 1952, **8**, 82–85.
15. Ramond, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
16. Shoben, E. J., Jr. The Wechsler-Bellevue in the detection of anxiety: a test of the Rashkis-Welsh hypothesis. *J. consult. Psychol.*, 1950, **14**, 40–45.
17. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
18. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
19. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
20. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
21. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–66.
22. Welsh, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65–72.
23. Westrope, Martha R. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.*, 1953, **48**, 515–524.
24. Winne, J. F. A scale of neuroticism: an adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117–122.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Personality Correlates of Certain Religious Attitudes[1]

William E. Broen, Jr.
University of Minnesota

A sample of 140 university freshman males was selected from liberal arts college class lists and divided into three groups representing three different degrees of "religiosity." The criterion for the separation of these groups was a Religiosity Index defined as the sum of an individual's standard scores on three Thurstone religious attitude scales: Attitude Toward the Bible, Attitude Toward God, and Attitude Toward Sunday Observance.

Various personality characteristics of these three groups were compared using the MMPI, and Welsh's Anxiety Index and Internalization Ratio. The only significant difference was between the religious and nonreligious groups on the MMPI *Pa* scale with the religious group scoring higher.

It was felt that the largely negative results might have been due to an incorrect grouping of persons on an impure "religiosity" continuum, when the personality correlates of possible differences in emphasis upon the various aspects of religion should have been investigated separately. Therefore, the three attitude scales, each of which was felt to tap somewhat different aspects of religious belief, were correlated separately with the *L* scale and the ten clinical scales of the MMPI and with Welsh's AI. Significant correlations of $-.24$ between the Attitude Toward the Bible scale and MMPI *D*, and $+.20$ between the Attitude Toward Sunday Observance scale and MMPI *Pt* were found. Both correlations were in the hypothesized direction.

Corroborative evidence for the statement that the several aspects of religion are stressed differentially in the lives of various persons was found in the intercorrelations of the attitude scales. The Attitude Toward the Bible scale and the Attitude Toward God scale (both with a positive belief content) were found to correlate with each other to a much greater degree than either did with the Attitude Toward Sunday Observance Scale with its "thou shalt not" content ($r = +.75$ vs. $+.49$ and $+.44$ respectively). The existence of separate "positive" and "negative" religious factors is therefore hypothesized.

These data suggest that factor-analytic research in this area might prove fruitful, with the purification of variables leading to more conclusive findings.

Brief Report
Received October 19, 1954.

---

[1] An extended report of this study may be obtained without charge from William E. Broen, Jr., 524 14th Ave. S. E., Minneapolis 14, Minnesota, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4437 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

The Use of Demographic Characteristics in Predicting Response to Hospitalization for Tuberculosis[1]

Louis J. Moran, George W. Fairweather, Robert B. Morton,
and Laurence S. McGaughran[2]

VA Hospital, Houston, Texas

Every year, almost half of all patients in tuberculosis hospitals leave the hospital against medical advice (3, 4). The prevalence of such discharges constitutes the greatest single obstacle to the control of tuberculosis in this country. Indeed, a recent investigation (3) indicates that the effectiveness of medical treatment for tuberculosis is now such that the major deterrent to success lies not in remedial techniques but in the inability or unwillingness of patients to stay for the necessary hospitalization period.

The psychological problems of tuberculous patients have been the subject of much study, and a staggering amount of material has been published on "the psychology of the tubercular." The reader is referred to Barker, Wright, Meyerson, and Gonick (1) for an exhaustive and critical 1953 review.

With few exceptions these studies may be grouped roughly into three categories. The majority are classified by Barker *et al.* as "expert opinion." Typically, these are observations on psychiatric, psychological, or social aspects of tuberculosis made by various specialists in their clinical practice. Another large group might be called "*ad hoc* explanation" studies. These usually categorize by frequency the probable reasons for quitting hospital treatment. An excellent review of these studies is given by Tollen (4). A third group might be categorized as "personality determinism" studies. Usually these involve the application of psychological tests or systematic interviews. Some of the studies in this third group seek to describe general "personality traits of the tubercular." Others attempt to compare the personality traits of patients who leave against medical advice with those of patients who remain for definitive treatment.

Problems of Method

Without an attempt here to analyze the approaches in detail, certain general questions concerning methodology may be raised. These methodological problems are by no means specific to studies of the conditions under which some tuberculous patients quit the hospital and others stay. However, there seems to be a particular need in this area for a closer examination of (*a*) criteria for selection of subjects, (*b*) sampling techniques, and (*c*) isolation of relevant variables.

Criterion. All studies designed to discriminate the characteristics of people who will remain for complete treatment from those who will leave before completion of treatment have used the administrative designation, "regular" (Maximum Hospital Benefit) vs. "irregular" (Against Medical Advice, AWOL, or Disciplinary) discharge as criterion for selection of subjects.

---

[1] We wish to acknowledge our indebtedness to Professor Trenton W. Wann, department of psychology, Rice Institute, and part-time staff member, Houston VA Hospital, for his contributions to the original formulation of the research project and for frequent consultations on theoretical issues during this study. We also wish to acknowledge the assistance given by James Hummell, department of mathematics, Rice Institute, in statistical consultations. Much credit is due to Drs. Daniel E. Jenkins, Irving Chofnas, and Hollis G. Boren, tuberculosis specialists, whose guidance and support made this study possible.

[2] Professor L. S. McGaughran, department of psychology, University of Houston, is a part-time staff member, Houston VA Hospital.

The conditions under which either of these administrative labels is applied are so varied, however, that they have little meaning as experimental criteria. For example, in the hospital under study here, 186 regular discharges were given in the period 1949–1952. Of these regular discharge patients, 14 per cent were found not to have tuberculosis at all and were quickly discharged, 18 per cent were found to have inactive disease and discharged, 12 per cent left "at their own risk," 7 per cent were discharged while still ill after emergency treatment, and only 49 per cent entered with active disease and remained in the hospital to follow a definitive treatment program.

Hence, an administrative regular discharge may include some people discharged without treatment, some with still active disease, some treated until well, and even some who leave at own risk, which is behaviorally equivalent to leaving "against medical advice." Although the irregular discharge category is relatively more homogeneous, it encompasses reactions other than leaving against medical advice, e.g., discharge for disciplinary reasons.

The heterogeneity and overlap of groups separated in terms of this administrative criterion militate against the discovery of psychological dimensions that might differentiate the two groups. At least, the complexity of such a criterion greatly complicates attempts to predict it. The development of a simpler definition of final response to hospitalization seems necessary.

Sampling. In most studies of irregular discharge, sampling has been postcriterion. That is, subjects are selected for study after discharge. In some studies, a cross section of patients in the hospital has been tested and the data analyzed after the patients have met the criteria of regular or irregular discharge. Other studies combine these sampling procedures.

In an area where so little factual information is available, precriterion sampling seems especially important. Actual behavior in the hospital may then be observed directly. The specific conditions under which the behavior occurs are accessible for objective measurements. And, most important, repeated measurements may be taken during the course of hospitalization.

If samples are selected for intensive study from currently hospitalized patients, a device for stratifying samples in terms of probable final discharge type would greatly enhance the efficiency of the design. It would eliminate the time period, often years, necessary for patients to reach the final discharge criterion. Comparison groups might be equated for size. Patients who would not eventually meet criterion could be excluded beforehand. The matching or equating of groups on important variables would be simplified. And, most important, evaluation of the effects of experimental intervention would be greatly facilitated, e.g., the reduction of within-group variance by the use of stratified samples.

At present, no technique is available for determining upon admission to the hospital the probability of regular or irregular discharge. The development of such a technique for stratifying samples would contribute in a number of ways to intensive studies of the process of adaptation to hospitalization.

Isolation of variables. The usual experimental approach to the problem of irregular discharge relies on psychological tests or interviews to detect differences in personality traits between regular and irregular discharge patients. The assumption implicit in these studies is that personality traits, "resident" in the individual, constitute the crucial determinants of his ultimate response (stay or quit) to hospitalization.

It seems equally reasonable to assume that other factors, entirely independent of personality, may be just as important correlates of such behavior as the personality that the patient "brings with him" into his particular hospital environment. For example, two people with hypothetically identical personality traits but with different family constellations, different incomes, different stages of disease, of different races, etc. might follow entirely divergent paths insofar as staying in or quitting the hospital is concerned. Furthermore, patients in the same hospital may be adjusting to quite different objective situations. Depending on the physical condition of the patient, treatment may range from simple rest to repeated strenuous procedures (such as months of periodic insertion of air into the peritoneal cavity) to major, disfiguring surgery. A patient may be painfully ill or asymptomatic, comfortably pensioned or ruined financially, rooming with a friendly or unfriendly patient, etc.

The contribution of these extra-personality factors to the variability of response to hospitalization might well obscure much of the variability contributed by personality traits if the latter are treated independently. Isolation and systematization of objective situational differences within the hospital and other such extra-personality variables, in a form suitable for controlled experimentation, are a logical prerequisite to studying the conditions under which some patients stay for complete treatment and others leave against medical advice.

The Problem

The aims of the present study are: (*a*) specification of simple criteria for final response to hospitalization, and (*b*) the development of a device for selecting samples, on admission to the hospital, with a predicted final response (stay or quit) to hospitalization.

Method

The criterion for selection of subjects is based on a simple stay-response or quit-response dichotomy. Stay response: the patient entered with active disease and stayed until discharged by his physician as having received adequate treatment. Quit response: the patient entered with active disease and left the hospital against the advice of his physician. Both groups entered ill, both groups began a treatment program, one group stuck it out, the other quit in midstream. The present study seeks to predict only one or the other of these two final responses.

Note that this selection excludes regular discharge patients who left after diagnosis only, after emergency treatment only, or at their own risk. This selection eliminates 51 per cent of the "regular discharge" sample, and also excludes disciplinary discharges and transferred or expired cases.

Data for this study were derived from the clinical record folders of discharged patients. The clinical record folder of each patient in a Veterans Administration hospital contains an account of his medical treatment history and certain demographic data, e.g., age, race, occupation. Forty such items of information available in the records were selected to serve several purposes, one of which involves prediction of final response to hospitalization. Information on these items was abstracted from the clinical records of all patients discharged from the Tuberculosis Service of a large General Medical and Surgical (GM&S) Veterans Administration hospital from March 1949 through December 1952.

Of this total sample, all cases failing to meet the criteria of stay response or quit response and all cases with more than two "unknown" information items were excluded. This left a total sample of 268 discharges.

Only 22 of the 40 information items abstracted from the clinical records are known at the time the patient is admitted to the hospital. Information on these 22 items was tallied for the selected stay and quit groups.[3] Cutting points were determined by inspection and the significance of differences between groups evaluated by chi square.

Results

Of the 22 items, 9 served to differentiate significantly the groups. Items significant at least beyond the .05 level were: age on admission, service-connected disability pension, duration of illness, chronic vs. fresh disease, number of previous hospitalizations, type of any previous hospital discharge, permanent address, psychiatric diagnosis in addition to tuberculosis, and occupation.

The following items did not reach the .05 level of significance: month of admission, day of admission, war service, marital status, number of children, number of siblings, various combinations of diagnoses (other than neuropsychiatric), time in service, religion, race, stage of tuberculosis (minimal, moderate, far advanced), and the number of readmissions to this hospital.

Each of the nine significant information items yielded individual probability statements concerning the quit response, i.e., the proportion of quit responses to row totals in the chi-square tables. However, since these items did not represent independent events,[4] computation of the probability of quit response by direct combination of these individual probabilities was not possible. Nor, with a sample of this size, was it possible to compute a probability statement that took account of interdependent events. Therefore, no attempt was made to derive mathematically the probability of quit response by means of combining the individual probabilities of the nine items. Instead, each patient was assigned a score, derived from the nine items.

A system of weighted scores was selected in preference to a system which assigned a patient an unweighted score from 0 to 9, depending on the number of high probability items he "hit." On the original sample, the unweighted scores yielded a contingency co-

---

[3] When the item of information does not vary with time, e.g., race, all cases with more than one discharge from the hospital are eliminated. This is to prevent any one case from unduly increasing frequencies in one or perhaps both groups. If the information does vary with time, e.g., number of previous hospitalizations, each discharge is counted. Hence, for some items the sampling unit is "people," and for others "discharges."

[4] Let A represent the event of a quit response and $H_1, H_2, \ldots H_9$ the nine known hypotheses. The combined probability of any two hypotheses is:

$$P(A)\left[\left(\frac{P(A|H_1)}{P(A)}\right)\left(\frac{P(A|H_2)}{P(A)}\right)\right].$$

When this formula is applied to various pairs of hypotheses as occurring in the problem under consideration, it is found that the formula gives a value greater than one in almost every instance. Hence, any hypothesis of independence must be rejected.

efficient with the criterion of .26, $p < .001$, with the median as cutting point, while the weighted scores yielded a contingency coefficient of .47, $p < .001$.

Weighted scores were determined in the following manner. The log was ascertained for every probability value as derived from the nine significant information items. The score of a patient was the sum of the nine log values that applied to him. This arbitrary procedure served only to accentuate the direction of individual probability statements and to disperse scores. The score merely places the subject somewhere on a scale and is not itself a probability statement. However, the probability of quit response for points on the scale may then be determined.

Scores thus computed for all cases in the sample ranged, with intervals of .1, from 12.3 to 17.2. Below an arbitrary cutting point of 15.7 were 77 cases. Of these, 89 per cent stayed for definitive treatment (stay response). Above a cutting point of 16.5 were 57 cases. Of these, 90 per cent left against medical advice (quit response). This index served, then, to select from the total group sizable samples with a high probability of either stay response or quit response.

Prediction for this group of 1949–1952 discharges was spuriously high, of course, since the index was constructed from the very cases being predicted. Therefore, the index was cross validated on a sample of 1953 discharges.

Cross Validation

The first 103 successive discharges in 1953 who met the definition of stay response or quit response were selected, and the same scoring procedures and cutoff points employed. Below the cutting point of 15.7 were 21 cases, with 81 per cent stay response. Above 16.5 were 32 cases, with 72 per cent quit response.

Thus, the application of this index in 1953 would have served to isolate from this sample of 103 cases a group (21 per cent) with 8 chances in 10 of stay response and a group (32 per cent) with 7 chances in 10 of quit response.[5]

Table 1

Probabilities of Quit Response Given by Each of the Items in the Refined Form of the Index

Item	N	Chi square p	Cutting points	Probability of quit response
Service-connected disability	360	.001	yes	.75
			no	.53*
History of Tbc	348	.001	first onset	.51*
			chronic case	.75
Duration of illness	332	.001	<2 months	.41
			>2 mo. <2 yr.	.60*
			>2 yr. <5 yr.	.64
			>5 yr.	.79
Previous hospitalization	340	.001	none	.43*
			1 or 2	.63
			>2	.85
Previous discharges	332	.001	none	.57*
			all regular	.11
			all irregular	.88
			mixed	.64
Residence	363	.05	same city	.57*
			other	.68
Previous psychiatric diagnoses	352	.001	no	.58*
			yes	.83
Occupation	333	.01	unemployed	.83
			laborer	.64
			student	.27
			other physical	.62*
			other mental	.51

Note.—N differs for each item because occasionally information on one item is "unknown" for a particular case. The score given in such instances is the log of the modal p value of that item, which is indicated here by an asterisk.

No predictions were made for cases that fell between the scores 15.7 and 16.5 (49 per cent of sample). It should be noted, however, that the index was devised not to predict all cases on admission but to select homogeneous samples from the total group. When used in this way, with cutoff points of 15.7 and 16.5, the index yielded a contingency coefficient of .48, $p < .001$, on the cross-validation sample.[6]

---

[5] It should be noted that these are *conditional predictions*. Predictions on current cases must be made only on cases with diagnosed active tuberculosis. This eliminates the miscellany of administrative regular discharges that are not classified here as stay response. Also, the small proportion of cases that expire, transfer to another hospital, or receive a disciplinary discharge must be dropped from a predicted sample.

[6] All contingency coefficients reported in this study are derived from 2×2 tables; hence, the maximum possible C is .707.

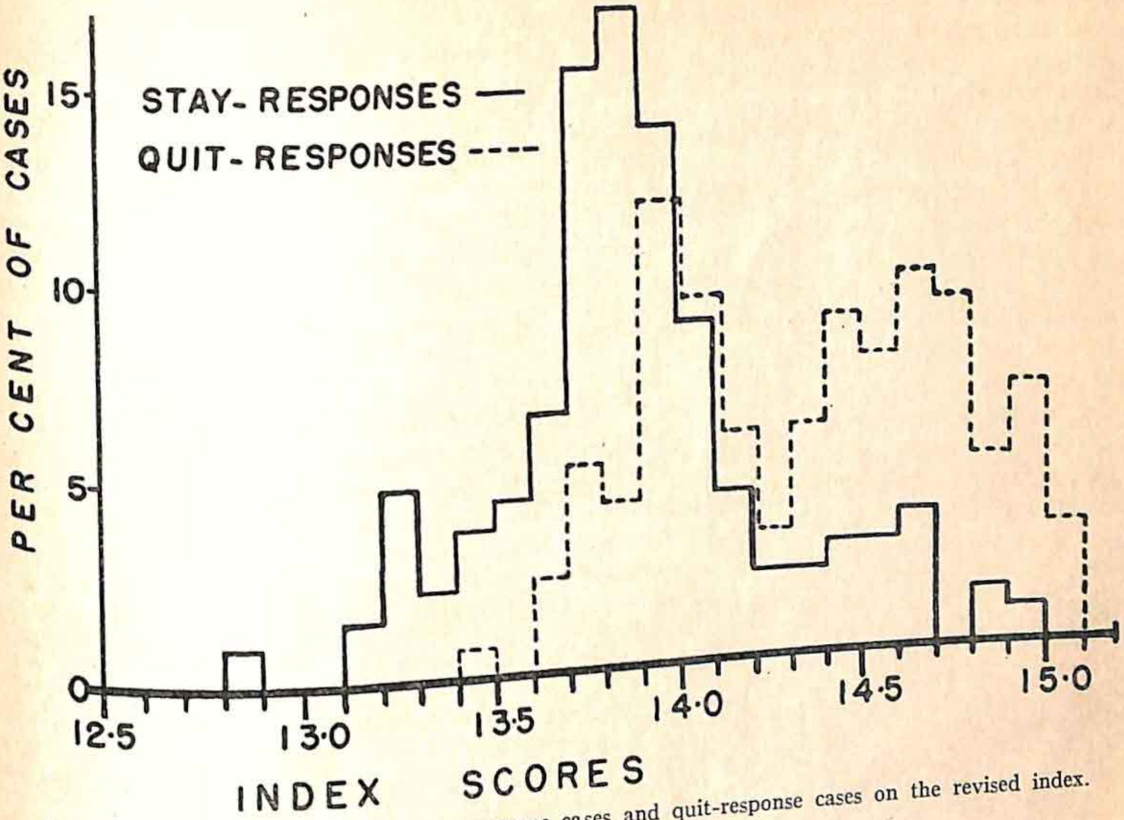

Fig. 1. Distribution of scores for stay-response cases and quit-response cases on the revised index.

The Revised Index

The final form of the index was achieved by combining the discharges in all five years, 1949–1953. On cross validation the discrimination power of all but the "age" item remained significant. Scores on the final form were derived from the remaining eight items exactly as in the original form. The individual probabilities furnished by each item are given in Table 1.

The scores of each group distribute as shown in Figure 1.

The validity of this index, i.e., the correlation between scores and stay or quit responses, clearly depends on the cutting points selected. If the index were used to attempt predictions on every case, the median (14.08) might be selected as the cutting point. The contingency coefficient would then be .40, $p < .001$. Other cutting points may be selected that yield higher validities. For example, cutting points of < 13.9 and > 14.1 isolate 79 per cent of all cases and yield a $C = .50$, $p < .001$. If homogeneity rather than size of sample is emphasized, cutting points of < 13.7 and > 14.5 isolate 40 per cent of all cases with $C = .59$, $p < .001$. The last cutting points, for example, isolate 39 cases with 8.5 chances in 10 of stay response and 103 cases with 8.8 chances in 10 of quit response.

Discussion

By means of this index, samples of a given size and with a given probability of stay or quit response may be selected from the population entering the hospital under study. This fulfills the aim of the present study.

Generalization of these results to other tuberculosis hospitals must, of course, await cross-validation studies. However, a recent unpublished dissertation by Speth (2), conducted independently in an eastern Veterans Administration hospital, suggests some likelihood of the generalized nature of these relationships in Veterans Administration hospitals. In conjunction with personality measures, Speth also collected certain demographic data in an effort to differentiate regular discharge from irregular discharge tuberculous patients. As in the present study, age, race, religion, number of dependents, and marital status were found not significant. Also in agreement with the present findings, service connection, length of illness, and previous irregular discharge were found to differentiate the groups.

The immediate value of this index is seen as a device for the economical preselection of stratified samples for intensive study during the course of hospitalization. It also may be applied to advantage in studies of irregular discharge that are designed to evaluate the effects of experimental intervention, in that within-group variance may be reduced by using groups stratified in terms of probable final discharge.

The predictive efficiency of these simple demographic data serves to emphasize the importance of variables other than personality traits as correlates of final response to hospitalization. As noted earlier, objective situations, both within and without the hospital, may vary greatly from patient to patient. Systematization of these variables as well as personality variables seems a logical preliminary to the study of adaptability to prolonged hospitalization.

Summary

About half of all hospitalized tuberculous patients leave the hospital against medical advice. An overview of literature on this problem suggests the need for a closer examination of (*a*) criterion for selection of subjects, (*b*) techniques of sampling, and (*c*) isolation of relevant extra-personality variables. A criterion for final response to hospitalization is specified and an index derived from demographic data that may be used to predict the final response of select samples. On cross validation, this index yields a C of .48, $p < .001$, between score and final response for the predicted samples. A refined form of this index is then described and its application as a technique for stratifying samples for intensive study or for evaluating effects of experimental intervention is discussed.

Received May 11, 1954.

References

1. Barker, R. G., Wright, Beatrice A., Meyerson, L., & Gonick, Mollie R. *Adjustment to physical handicap and illness: a survey of the social psychology of physique and disability.* New York: Social Science Research Council, 1953. (Bulletin 55 revised.)
2. Speth, E. W. An investigation of emotional and social concomitants in the irregular discharge rates of hospitalized patients with tuberculosis. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1953.
3. Tempel, C. W., *et al.* An analysis of hospital records of patients discharged from a large tuberculosis service. *U. S. Armed Forces Med. J.*, 1953, 4, 1719–1733.
4. Tollen, W. B. Irregular discharge: the problem of hospitalization of the tuberculous. *VA Pamphlet 10–27*, 1948.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

Results of Treatment of Enuresis by a Conditioned Response Method

Barclay Martin and Delores Kubly
University of Wisconsin

Several investigators have reported using conditioning techniques with great success in the treatment of enuresis. Mowrer and Mowrer (3) treated 30 children between three and thirteen years of age by this method and reported that all were cured of enuresis within four to eight weeks. Morgan and Witmer (2) successfully cured four out of five chronic enuretics. The fifth was discovered to have bladder trouble that required medical treatment. Davidson and Douglass (1) treated 20 chronic enuretics, ages six to thirteen, and reported 15 to be cured and 5 to be greatly improved (not more than one wet bed per week).

These studies certainly suggest that this procedure is very efficacious in stopping enuresis. The samples used, however, were relatively small, and there is a need for additional results on larger numbers of children before the general effectiveness of the method can be assessed. It was toward this end that the following data were obtained.

Procedure

The information was obtained by means of a questionnaire filled out by parents whose children had been treated for enuresis at home with a conditioning apparatus. Questionnaires were sent to all people who had bought this commercially sold apparatus between January 1951 and July 1953.[1] Respondents were asked not to put their names on the questionnaire in order to assure anonymity.

The apparatus itself consists of two wire mesh "foil pads" separated by a piece of cloth. When the circuit between the foil pads is completed as a result of urination, a buzzer sounds and the child is awakened. In terms of conditioning theory, bladder tension is the conditioned stimulus, the buzzer is the unconditioned stimulus, and waking up is the response that is to be associated with bladder tension. As time goes by, it is also hoped that the somewhat more temporally removed response of sphincter control will become associated with bladder tension, thus allowing the child to sleep through the night dry.

Parents who buy this equipment are requested to consult a physician before using it in order to rule out organic factors.

Results

One hundred and eighteen questionnaires were returned out of the 220 that had been sent out. Part of the information thus obtained is tabulated in Table 1. The four points on the improvement rating scale are defined as follows:

Cure: Three recurrences or less since the end of treatment.

Marked improvement: One to two recurrences per month since the end of treatment.

Temporary improvement: Two to six recurrences per month, or intermittent dry and wet periods, or relapse and reusing apparatus.

Failure: Unqualified negative reply.

The more important results can be summarized as follows:

1. When the Cure and Marked Improvement categories are combined, the treatment is found to be relatively successful in 74% of the reported cases. This figure must be accepted with some caution since it is not unreasonable to assume that among the 102

---

[1] The authors wish to express their appreciation to the Health Guardian Corporation for its cooperation in this study.

Table 1

Results of Enuresis Treatment

	Improvement rating									
	Cure		Marked improvement		Temporary improvement		Failure		Total	
Group	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Both sexes	66*	56	21	18	14*	12	17	14	118*	100
Males	42	53	13	17	11	14	13	16	79	100
Females	21	60	8	23	2	6	4	11	35	100

* The *N* of 118 includes four *S*s for whom sex was not reported.

cases who did not return the questionnaire, there may be a higher percentage of negative results. Respondents, however, were told in an accompanying letter that the investigators were just as interested in negative as positive results.

2. There is a significant difference between the sexes in terms of the occurrence of enuresis. Boys make up 69% of the 114 cases in which the sex was reported, and the probability of a chance deviation from 50% is less than .01. This finding is consistent with previous findings (1, 4).

Seventy per cent of the boys were in the Cure or Marked Improvement categories as compared with 83% of the girls, but the sex difference was not significant. If only those cases in the Cure category which report no recurrences are considered, the following is found: No recurrences are reported by 18% of the boys as compared with 40% of the girls. This difference reaches the .02 level of significance.

3. The average age of the children was 9.3 years with no significant sex differences. The ages ranged from 3.5 to 18.5 years. Improvement was not associated with age.

4. The average length of treatment was about six weeks. For boys the average was 6.7 weeks and for girls the average was 4.1 weeks, with a difference that reaches the .02 level of significance. The greater frequency of enuresis in boys, and the tendency for girls to show a higher percentage of cures without recurrence and to be treated for a shorter period of time, suggest that in general girls are less susceptible to enuresis and are somewhat more easily cured.

5. For the cases in the Cure and Marked Improvement categories, an average time of 14.5 months had elapsed since termination of treatment. There were no significant sex differences.

6. The children had had enuresis for an average of six years before this treatment. There were no significant sex differences. Eighty-seven per cent of the children were reported to have been bed wetters since birth. The children were considered to have had enuresis only after the age of three.

Discussion

These results must be considered as rough approximations. The questionnaire approach prevents one from applying exactly similar procedures to all subjects and, in addition, one must rely upon the memory of the parents. Also, there is no control group of enuretics that would take into account "spontaneous" cures. In spite of these shortcomings, the results indicate that this procedure is effective in stopping enuresis in a substantial number of cases.

Conditioning or habit training procedures are frequently referred to by clinicians as being merely "symptomatic," with the implication that really successful and lasting cures for symptoms such as enuresis can only be achieved by more intensive therapeutic attention to underlying personality disturbances. In the case of enuresis the present writers would like to suggest the following viewpoint.

There is no sharp dichotomy between children for whom enuresis is a rather isolated habit in an otherwise well-adjusted personality and children for whom enuresis is one manifestation of deep-seated anxiety and seriously disturbed interpersonal relations, usually in the family context. The alternatives probably represent extremes on a dimension. The greater the investment that the child has in the enuresis as an interpersonal expression—for example, as a passive way of expressing hostility and defiance toward the parents—the less likely is it that the habit training procedures will be effective. However, the results of the study suggest that a substantial number of enuretics are not so emotionally disturbed as to render these procedures ineffective.

Of course, it would be wise for parents to obtain a psychological evaluation of the child as well as a physical examination before using the conditioning apparatus.

Summary

The results of treating enuresis by a conditioned response method are reported in this study. It was found from questionnaire responses by parents of 118 children that this procedure was relatively successful in 74% of the cases. Girls were found to have enuresis less frequently, and to be somewhat more easily cured when they did have it.

Received May 24, 1954.

References

1. Davidson, J. R., & Douglass, E. Nocturnal enuresis: a special approach to treatment. *British med. J.*, 1950, **1**, 1345–1347.
2. Morgan, J. J. B., & Witmer, F. J. The treatment of enuresis by the conditioned reaction technique. *J. genet. Psychol.*, 1939, **55**, 59–65.
3. Mowrer, O. H., & Mowrer, W. M. Enuresis—a method for its study and treatment. *Amer. J. Orthopsychiat.*, 1938, **8**, 436–459.
4. Sheldon, W. Discussion on enuresis. *Proc. roy. Soc. Med.*, 1944, **37**, 348–349.

Journal of Consulting Psychology
Vol. 19, No. 1, 1955

MMPI Responses of Patients with Multiple Sclerosis[1]

Franklin C. Shontz
Highland View Hospital, Cleveland, Ohio

In the present research, the Minnesota Multiphasic Personality Inventory was administered to a group of sixteen patients hospitalized for multiple sclerosis and to a group of sixteen patients hospitalized at the same institution for chronic illnesses other than multiple sclerosis. The groups were matched (by groups) for chronological age, and the average MMPI profiles for both samples were compared with each other and with the results reported by Canter (1) in a previous research.

Significant differences were found between the three samples (Canter's "MS" group, and the present "MS" and "Other Chronic Illnesses" groups) only with respect to the *Hs*, *Hy*, *D*, and (?) variables. On both the *Hs* and *Hy* scales, the three groups could be ranked in descending order as follows:

1. MS group from Canter's research.
2. MS group from present research.
3. Other Chronic Illnesses group from present research.

All differences were significant at a probability level of .05 or less. On the *D* scale, the present group of MS patients was again found to obtain a significantly lower mean score than Canter's subjects, but this lower value was not significantly different from that of the Other Chronic Illnesses group in this research. Finally, the present MS patients were also found to produce more (?) responses than the Other Chronic Illnesses group, although differences in the administration of this scale made cross-study comparison impossible.

In short, the present sample of patients with multiple sclerosis could be differentiated from a sample of patients with other chronic illnesses by the evidence of higher levels of self-concern, body anxiety, and uncertainty of response in the former group. Both samples obtained about equal *D* scale scores.

Despite the occurrence of certain differences, the average MMPI profile for the present MS group was found to be very similar in form to that obtained by Canter. The lower values which were obtained in this research on the *Hs*, *D* and *Hy* scales may probably be attributed to the higher mean age and longer duration of illness of the patients in the present study (implying more adaptation to the disease process itself) and to the fact that many of the patients used in this research have been permanently hospitalized for some time and therefore exist in a relatively secure and nonfrustrating environment.

The ambiguity of the disease process itself is held to account for the higher level of (?) responses for certain individuals in the presently tested group of patients with multiple sclerosis.

Brief Report
Received November 1, 1954.

Reference

1. Canter, A. H. MMPI profiles in multiple sclerosis. *J. consult. Psychol.*, 1951, **15**, 253–256.

---

[1] An extended report of this study may be obtained without charge from Franklin C. Shontz, Highland View Hospital, Cleveland 22, Ohio, or for a fee from the American Documentation Institute. To obtain it from the latter source order Document No. 4438 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

NEW BOOKS AND TESTS

Books

Arbuckle, Dugald S. *Student personnel services in higher education.* New York: McGraw-Hill, 1953. Pp. x + 352. $4.75.

In addition to the content usually found in a discussion of personnel services, such as admission, orientation, group activities, health, student aid, and housing, the author has included chapters on teaching and religious services. The book is addressed to all faculty members and is written in an inviting, nonpedantic style. The chapters on counseling and teaching present a good summary of the basic contemporary ideas about these processes in a student-centered framework. If a good percentage of our faculties would read and react to these ideas, a real contribution to higher education would be made. It is in the chapters on counseling and teaching that Arbuckle comes closest to solving a common problem in the field, namely, not to lose sight of the dynamic person when discussing services well planned for his development. The 74 pages of appendices, which include many blanks and specific illustrations of personnel activities in practice, will be helpful to personnel workers in the field. The reviewer feels a better coverage of the periodic literature could have been achieved without distracting from the readable and integrated nature of the volume.—*F. McK.*

Bellak, Leopold. *The Thematic Apperception Test and the Children's Apperception Test in clinical use.* New York: Grune & Stratton, 1954. Pp. x + 282. $6.75.

In this clinical manual for the TAT and CAT, Bellak elaborates his techniques for recording and interpreting which have been presented more briefly elsewhere (see *J. consult. Psychol.*, 1950, **14**, 161; 1952, **16**, 159; 1953, **17**, 156). Although he gives short summaries of other approaches, the author leans most heavily on his TAT blank, which is used in presenting the longer illustrative cases. An analogous form is used in the interpretation of the CAT protocols. References to completed research are minimal, but there is a short chapter outlining needed studies. Because the informal richness of the picture-story method is best revealed by studying many examples analyzed from varying viewpoints, this manual will be instructive for clinicians and students. It is probably not, in itself, a sufficient textbook in projective methods.—*L. F. S.*

Brand, Howard. (Ed.) *The study of personality: a book of readings.* New York: Wiley, 1954. Pp. xvi + 581. $6.00.

Brand's book of readings is a well-constructed instructional tool. The selections are diverse but all are characterized by a high degree of relevant readability; there is hardly a dull one in the lot. The editor has supplied integration by general introductions to the sections and by brief notes which precede each article. The book will be useful for instructors who like their students to read, think, and discuss. It will be less helpful to those who wish to present a systematically organized theory.—*L. F. S.*

Burrow, Trigant. *Science and man's behavior.* New York: Philosophical Library, 1953. Pp. xii + 564. $6.00.

After receiving training in both medicine and psychology, Burrow became interested in the psychoanalytic movement and participated in the introduction of psychoanalysis in America. Despite the individualistic emphasis of his training, he felt that an organismic point of view, which included a concern with the emergent characteristics of mankind as a social group, was necessary for the proper understanding of human behavior. Burrow presented his views by correspondence to a number of eminent scientists and the present volume is inspired by this correspondence. Unfortunately Burrow suffered a fatal illness in 1950, and the present volume is an integration and a completion of his work prepared by a close associate, William E. Galt. The volume, which cuts across many disciplines, shows a continuous concern with some of the destructive features of man's group behavior. In its effort to get

---

Note.—The reviews were prepared by the Editor and the Advisory Editors, who may be identified by their initials.

the reader to view mankind with a detachment which the reader's individual past scarcely permits, the work is most provocative. Although the reader is certain to be most sympathetic with Burrow's concern for the fate of man, and respectful of the scope of his scholarly allusions, he is likely to find the reading unusually difficult. In anticipation of this, a quotation from the author is most noteworthy: ". . . it is the automatic tendency of the student confronted with the principles of phylobiology to shy away from them, and in support of his irresolute policy to indict my formulations as being obscure and unacceptable. . . ." The author provides his irresolute reader with a ten-page glossary offering not only new words and new concepts, but also old words with new meanings and old meanings with new words.—*J. R. W.*

Eidelberg, Ludwig. *An outline of a comparative pathology of the neuroses.* New York: International Universities Press, 1954. Pp. 263. $4.50.

"This book is being written for advanced students of psychoanalysis, and should be read only by them." With this admonition, and with the introductory explanation that the author proposes in this volume a method for the "systematic, comparative examination of analytic material" as a substitute for quantitative measurement in psychoanalysis, one approaches the book with considerable interest and great expectations. Careful reading, however, leaves one wondering whether the title, the preface, and the introduction were not intended for another book. Unless there is much more than meets the eye, eight of the nine chapters deal rather simply, systematically, and briefly with Freudian psychology in strict accordance with the gospel of Freud. Even commonly used and accepted terms and concepts are defined in elementary fashion. The ninth chapter epitomizes even more succinctly the material summarized in the previous eight. Only in the last page or two is the "comparative" approach used. Were it not for the pretentious claims of the author at the outset, this book would be approached as an excellent summary of the field of psychoanalysis. Much of the value of the book lies in the great number of short illustrative samples culled from the obviously rich psychoanalytic experience of the author.—*M. K.*

Frank, Lawrence K.; Harrison, Ross; Hellersberg, Elisabeth; Machover, Karen; & Steiner, Meta. Personality development in adolescent girls. *Monogr. Soc. Res. Child Develpm.,* 1951, **16,** No. 53 (issued, 1953). Pp. 316 (paper).

This study of the personality of adolescent girls is executed primarily through the group application of the Thematic Apperception Test, Horn-Hellersberg Test, Human Figure Drawings, Rorschach test, and graphological analysis of the TAT records. One hundred subjects were drawn from each of the following age ranges: prepuberal, puberal, and adolescent. The collaborators provide a chapter for each test in which they include discussions of the method of testing and of the personality qualities implied by the results. Despite the idiographic emphasis of the total work, numerous group comparisons are included. The latter portions of the monograph include an interpretive summary of findings, suggestions for schools and agencies, and an appendix which includes a summary and discussion of the records of several individuals from each of the age groupings.—*J. R. W.*

Gruenberg, Sidonie Matsner. (Ed.) *The encyclopedia of child care and guidance.* Garden City, N. Y.: Doubleday, 1954. Pp. 1016. $7.50.

The Encyclopedia of Child Care and Guidance is a comprehensive encyclopedia covering the whole range of child growth and development. All areas of child guidance—physical, emotional, educational, cultural—are touched on. The emphasis is on practical handling of situations arising in daily living. It is well edited and has excellent cross indexing so that all information on a special topic is clearly indicated and easily found. The illustrations are attractive and add a light, human touch to the book. There are thirty chapters on such topics as "Family Life is Changing," "How Children Begin to Learn," "Infancy: Off to a Good Start," etc., all contributions by specialists in their areas. The articles are readable, simply written in nontechnical language, and reassuring to parents. The reviewer has mixed feelings. While the book is very suggestive and helpful, it will have all the advantages and disadvantages of a family medical book. Such a book is valuable to parents as a source of needed information, a book to which they can constantly go for better understanding of daily situations. But it is not so valuable if it increases "self diagnosis" and allays fears so that there is a delay in getting professional help when needed. *The Encyclopedia of Child Care and Guidance* should, however, prove useful for all those working with parents and families. It is comprehensive and complete in its coverage, and the practical suggestions should be helpful in meeting each day's problems.—*B. M. L.*

Jones, Marshall R. (Ed.) *Nebraska symposium on motivation: 1954.* Volume II of the series in Current Theory and Research in Motivation. Lincoln, Nebr.: Univer. of Nebraska Press, 1954. Pp. x + 322. Cloth, $3.50. Paper, $3.00.

This is the second University of Nebraska symposium on motivation held at the University on February 11 and 12, and April 1 and 2, 1954. I. E. Farber writes on anxiety as a drive state, John Atkinson on measuring motivation through sampling imaginative thought, Benbow Ritchie on the logical analysis of the laws of motivation, Leon Festinger on motivation and social behavior, George Klein on need and cognitive control, and Henry Nissen on drive as innate determinant of behavioral organization. The papers are scholarly and stimulating although highly diverse. No integration is achieved, but none is attempted. There is a lot to be said about motivation, and obviously a lot of it being said. The reading of it remains exciting.—*W. A. H.*

King, H. E. *Psychomotor aspects of mental disease.* Cambridge, Mass.: Harvard Univer. Press, 1954. Pp. xiv + 185. $3.50.

A research monograph, reporting the application of tests of reaction time, speed of tapping, and finger dexterity to patients of various degrees of disturbance and to normal subjects. Schizophrenics were retarded in all functions, in degrees significantly related to the severity of their disorders. The form of their practice learning curves did not, however, differ from that of normal persons. Psychoneurotics differed little from the normal controls either in level of performance or in learning functions. The data suggest that the schizophrenics are about as well differentiated by these simple motor tests as by many supposedly more subtle and dynamic techniques. The implications are thought provoking—are we striving too hard to be subtle and overlooking the obvious?—*L. F. S.*

Maslow, A. H. *Motivation and personality.* New York: Harper, 1954. Pp. xiv + 411. $4.50.

Clinical psychology, for which personality theory is such a central matter, could benefit from a book which effectively fuses the humanistic aims of a profession and the investigative and logical methods of a science. Toward this ambitious target Maslow directs his lance in his collection of interesting papers. He fails, but the failure is a spectacular one in the tradition of Don Quixote. While tilting fruitlessly with the windmill of "atomism," he suggests a number of novel ideas for thoughtful consideration and vigorous further research. While revealing either a startling misconception or a frank distaste for almost all that has passed so far for psychological science, he reaffirms the respect and unselfish concern for the dignity of others that are fundamental to the personal aspect of the clinician's labor. And if he misunderstands the role of theory in science, it does not keep him from engaging in some creative empirical work that perceptibility presses back the wall of ignorance even as it violates his abstract discussions of theory's nature. For its brief reports of intriguing investigations and as humanistic literature, the book is a delight. As an articulate integration of the values underlying clinical practice and the techniques of thought and inquiry that provide the knowledge necessary to implement these values, it represents much more *con*fusion than fusion.—*E. J. S.*

Piaget, Jean. *The construction of reality in the child.* New York: Basic Books, 1954. Pp. xiii + 386. $6.00.

This volume is one more testimonial to Piaget's care in making observations and to his ingenuity in organizing and interpreting them. He here raises the question of how the child comes to "know" his universe, in the sense of attaining concepts of objects, of space, of time, and of causality. For most of the studies, his subjects were his own children as they behaved during the first eighteen months of their lives. His observations of their behavior were based partly on simple but well-controlled stimulus situations (finding a hidden toy, for example), partly on spontaneous reactions (watching persons come and go, for example). Because of their painstaking and detailed character, these observations provide badly needed information regarding the stages of development in early infancy. One cannot help noting, however, the absence of statements concerning needs, motives, or social situations. To many contemporary child psychologists, accustomed to interpreting early development in dynamic terms, this may seem a serious shortcoming. To others, not committed to a particular theoretical view, Piaget's strict adherence to observable behavior may seem a refreshing change. No psychologist interested in behavior development can help being stimulated by the observations reported here. Many will also sense from Piaget's conceptualization of his data a need for rethinking some of our dynamic hypotheses regarding infancy.—*A. M. G.*

Rogers, Carl R., & Dymond, Rosalind F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. x + 447. $6.00.

Books about psychotherapy fall, by and large, into two classes: summations of clinical experience and general theoretical discussions. In either case, they vary considerably with respect to both sophistication and degree of relationship to evidence. This work, a product of the University of Chicago Counseling Center staff, is, therefore, without precedent; it is a whole book of *research* studies in the therapeutic process and its outcomes. Eleven chapters are devoted to reports of empirical investigations of a controlled sort, and two are concerned with searching case analyses, methodologically unique in their reliance on quantified techniques and objective data. For its boldness of conception, the enterprise and devotion to explicit knowledge that it reflects, and its uniqueness as the only volume concerned entirely with empirical inquiries into the events of psychotherapy, *Psychotherapy and Personality Change* commands respect and deserves a wide and attentive audience. Between the lines, however, it constitutes a casebook in the frustrations and difficulties attendant upon research in this field. Lost cases, spoiled records, and a multiplicity of wild variables lurk like bogeymen in the background, raising questions about the suitability of data and marring the precision of tests. Too, the techniques of data analysis will not prove as convincing or appropriate to many readers as they apparently did to the participants in this investigation. Finally, the edge of novelty seems to have worn off the Chicago group's contributions. The studies reported here bear a disappointing degree of similarity to those published previously from this fertile source. One wonders if the concepts and basic ideas of client-centered counseling are not in need of a revision now that they have served their stimulating purpose. But no one else has done anything to challenge so wholesomely our thinking about therapy, to provide sounder information about what happens in the consulting room, or to offer a more constructive model for those who would replace intuition with communicable knowledge.—*E. J. S.*

Stanton, Alfred H., & Schwartz, Morris S. *The mental hospital.* New York: Basic Books, 1954. Pp. xx + 492. $7.50.

In this volume, social psychology takes a penetrating look at the structure and functioning of a mental hospital and comes forth with some insightful conclusions. The hospital, like other social institutions, has both a formal structure shown by its organization chart and an informal structure through which communications really pass. In many respects, the problems resemble those of a business or a university—the perceptions and misperceptions of "the front office" by the lower echelons of staff provide an example. Other issues are unique to the hospital. The study considers the patients as active participants in the social organization, not merely as passive recipients of the staff's efforts. Of special interest is the evidence of an intimate relationship between the staff's morale and the patients' behavior. The study deserves the attention of both clinical and social psychologists.—*L. F. S.*

Taylor, William S. *Notehand for psychologists.* (3rd Ed.) Northampton, Mass. (15 Pierce Hall, Smith College): Author, 1954. Pp. 5 (mimeographed). 21¢.

Syss V shorthnd wr devpd as early as t 1st century V our era; & syss V notehnd, whi is ls techcl, mu earlr. Bo syss save ti & energy. Few ps, howev, k shorthnd; & any shorthnd n dn b an xpert cn b rd o b t id. who wrote it & sn bcms "cold" f hm. Availbl now is a sys V notehnd whi is easly lrnd. It is designd primarly f pen or pencl notes, drafts V ltrs & MSS, etc., yt mst V it cn b typd. It reduces t ⋇ V chars i ordinary matl abo 30%, & psll matl, 35%. T sys is presentd i 5 mimeogrd pp. & cs. r availbl at approx cost. Send name, address, & 21¢ i stamps f ea c. to au.—*L. F. S. & W. S. T.*

Tests

Langmuir, Charles R. *Oral Directions Test.* Personnel Tests for Industry, ODT, Form S. Adults in industry. 1 form. (15) min. Record, 12 in., 33 r. p. m., or tape, 300 ft., 3.75 in./sec. ($12.00 with manual, key, 100 answer sheets); answer sheet ($4.00 per 100); manual, pp. 8 (35¢). New York: Psychological Corp., 1954.

A revised abbreviation of a test first published in 1946, the Oral Directions Test provides an appraisal of mental ability for industrial applicants that is relatively independent of reading skill. Examinees listen to a recording and mark an answer sheet to show their comprehension of directions. Corrected split-half reliability is about .87; the test correlates .85 with a conventional printed test. The manual gives preliminary evidence of discriminating above average and below average workers in industry, and provides tentative norms based mainly on textile and pharmaceutical employees.—*L. F. S.*

Manson, Morse P., & Ulett, George A. *The Rorschach Evalograph.* Record and analysis blank for Rorschach, pp. 24 ($3.50 per 10; $7.00 per 25; $25.00 per 100). Beverly Hills, Calif. (Box 775): Western Psychological Services, 1954.

Perhaps the most elaborate Rorschach blank so far devised, the *Evalograph* provides spaces for recording, scoring, and tabulating responses, color-printed miniature blots for locations, a summary and profile, and numerous aids to the examiner. It is an attractive and compulsively thorough product, a dream record to end all records, but one may properly question its economy. The recording and scoring can probably be done as effectively on less costly paper. The percentage table, the list of symbols, and the hints on scoring and interpretation belong in a manual, not in an expendable blank. The *Evalograph* is fun if you can afford it.—*L. F. S.*

JOURNAL OF CONSULTING PSYCHOLOGY

April, 1955

Vol. 19, No. 2

Contents

Correspondence Between Attitudes and Images of Parental Figures in TAT Stories and in Therapeutic Interviews: *Mortimer M. Meyer and Ruth S. Tolman* - - - - - - - 79

The Application of a Psychometric Scoring Procedure to a Group Modification of the Thematic Apperception Test (N-TAT): *Leonard D. Eron, Florence Sultan, and Frank Auld, Jr.* - - - - - - - - - - 83

The Effects of Sexual Stimulation Under Varying Conditions on TAT Sexual Responsiveness: *Paul H. Mussen and Alvin Scodel* - - - - - - - - - - - - - - - 90

The Diagnostic Significance of Usual and Unusual TAT Stories: *Jerome S. Miller and Alvin Scodel* - - - - - - - - - - - - - - - - 91

The Diagnostic Discrimination of the "Basic Rorschach Score": *Ben C. Finney* - - - - - 96

Psychotherapeutic Technique, Professional Affiliation, and Experience Level: *Hans H. Strupp* - - - - - - - - - - - - - - - - - - 97

Adjustment Changes in the Absence of Psychotherapy: *Rosalind F. Dymond* - - - - - 103

Language and the Interpretation of "Sexual" Symbolism: *Harold H. Mosak* - - - - - - 108

Conformance in Role Perception as a Personality Variable: *Theodore R. Sarbin and Curtis D. Hardyck* - - - - - - - - - - - - - - - 109

The Rorschach Experience Balance in Epileptics: *Paul W. Pruyser and Angela T. Folsom* - 112

A Note on Some Projected Familial Attitudes Associated with Rorschach Movement Responses: *Jerome L. Singer and Daniel A. Sugarman* - - - - - - - - - - - - - 117

Relationship Between Diagnostic Category and Deviant Verbalizations on the Rorschach: *William T. Powers and Roy M. Hamlin* - - - - - - - - - - - - - - - 120

The Measurement of Anxiety in Clinical Situations: *Arnold H. Buss, Morton Wiener, Ann Durkee, and Marc Baer* - - - - - - - - - 125

Cross Validation of the Organic Brain Pathology Scale for a Test of Symbol Arrangement: *Theodore C. Kahn* - - - - - - - - - - - - - - - - - - - 130

Digit Symbol Performance as a Function of Increasing Levels of Anxiety: *Joseph D. Matarazzo and Jeanne S. Phillips* - - - - - - - - - - - - - - - 131

The Discriminative Power of the Strauss-Lehtinen Battery in Adult Mental Patients: *A. A. Rafi* - 135

Normative Data on Progressive Matrices (1947): *Meredith W. Green and Josephine C. Ewert* 139

Relationships Between Verbal and Nonverbal Parts of the CTMM and WISC: *Grace T. Altus* - 143

Food Aversions: Some Additional Personality Correlates: *Wendell I. Smith, Elizabeth K. Powell, and Sherman Ross* - - - - - - - - - - - - - 145

The Effect of Pinhole Observation on the Flicker Fusion Frequency in Patients with Brain Pathology: *James C. Stauffacher, John B. Marks, and Albert F. Ax* - - - - - - - 150

Emotional Adjustment and Dieting Performance: *John Summerskill and C. Douglas Darling* 151

New Books and Tests - 154

Edn. Research
COLLEGE
Acc. No.

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 70 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 70 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 70-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. —— from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $—— for microfilm or $—— for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Extended report. The full report is prepared in the style specified by the *Publication Manual* (1), except that it may be typed with single spacing for economy in photoduplication by the ADI.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.*, 1952, 49, 389–449.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Correspondence Between Attitudes and Images of Parental Figures in TAT Stories and in Therapeutic Interviews[1]

Mortimer M. Meyer and Ruth S. Tolman
VA Regional Office Mental Hygiene Clinic, Los Angeles

The assumption has been made that in the TAT, by the construction of stories around ambiguous pictures, the individual organizes material from his own experience. The expression of consciously recognized autobiographical material in TAT stories, or, as Murray (8) puts it, the narration of "frank and unabashed autobiographies," has been the subject of several investigations. One of the earliest was Murray's (9) comparison of the results of the TAT with information obtained through other devices, among them interviews and autobiographies. His results were never subjected to statistical treatment, but clearly Murray regarded them as satisfactory validation. Harrison (5) used the method of matching reports based on the TAT with case history material of the same individuals and met with more than chance success. In another experiment Harrison found that biographical information could be deduced from the stories of mentally ill patients with a fairly high degree of validity. Rotter (10) on the basis of his investigation indicates his finding that "the identified figure" (the hero) has a history similar to the subject's own.

Markmann's (6) study of the relationship between stories told by 40 individuals in response to three of the TAT pictures and their past history revealed a more faithful reflection of this history in some pictures than in others. She found that stories to Picture 1 most frequently reflected the actual history, especially if the theme of the story was an aggressive one. Tomkins (12), in reviewing Markmann's findings, suggests that much work needs to be done to explore the conditions under which past history is reflected in the TAT.

Combs (4) in a study of TAT stories and personal experience as recorded in autobiographies found that approximately 30 per cent of TAT stories contained significant material drawn from the subject's own life experience. In terms of his criterion, Picture 2 encouraged the most transfer of personal experience.

Now in all these studies, and in others reported by Bell (1, pp. 207–252), there is material on the subject's attitudes toward or experiences with the parents or with parental figures, but little or nothing in regard to the attitudes which the subject attributed to these figures. In this same way, Bellak's (2) outline for analysis or interpretation of the TAT includes a section on "Attitudes to Superior (Parental) Figures or to Society"; but again these refer to the subject's attitudes, not to those ascribed to parental figures. An unpublished Ph.D. dissertation by Butler (3), "Parent Figures in TAT Records of Children in Disparate Family Situations," may deal with some aspects of this question for children, but does not touch on the situation for adults.

There are two recent studies, Mitchell's (7) and Singer's (11), which do involve attitudes ascribed to parents. Both use the TAT and both assume that the TAT stories correctly represent the subject's feelings about his parents' attitudes toward him. These studies are concerned primarily with familial attitudes expressed in stories to TAT cards in relation to socioeconomic status.

---

[1] Presented at the meeting of Western Psychological Association, Long Beach, California, May 20, 1954.

Hypotheses

It was the purpose of this study to explore specifically the correspondence between characteristics attributed to parents in TAT stories and in therapeutic interviews. The characteristics investigated were of two kinds: (*a*) attitudes toward the subject or patient which were ascribed to the parents; and (*b*) general descriptive images or pictures in which the subject has depicted the parental figures.

Using TAT stories and the content of therapeutic interviews of patients in a mental hygiene clinic, we attempted to test three hypotheses:

1. Patients who tell TAT stories involving attitudes ascribed to parents or images of parental figures will also in therapeutic interviews discuss attitudes which they ascribe to their own parents or images in which they are pictured. The converse also will be true. Hence, the TAT will furnish a basis for prediction.

2. Attitudes ascribed by patients to parental figures in TAT pictures will resemble those ascribed to their own parents (or parent substitutes) in therapeutic interviews conducted within a short period of time from the date of testing.

3. The kind of images in which parental figures in TAT pictures are described by patients will resemble images used in description of their own parents during these same therapeutic interviews.

It must be emphasized that in this study we were undertaking to discern only consciously expressed attitudes ascribed to the parents. No attempt was made to be subtle or interpretative in the notation of attitudes. Hence, if a description of a parental figure as "comforting and giving" indicated a concealed view of the parent as the exact opposite, this circumstance would not find inclusion in our data. Thus, possibly important aspects of TAT material such as reversals of conscious fantasy would not be revealed.

Method

The materials examined in this study came from the files of 50 patients in the Mental Hygiene Clinic of the Veterans Administration in Los Angeles. These materials were:

1. TAT stories in response to Cards 1, 2, 6 BM, 7 BM, 7 GF, and 18 GF—those cards in which figures appear which are the most easily identified or interpreted as parental ones; and

2. Records of ten therapeutic interviews which in most instances were the first ten interviews of therapy.

In order to avoid biased sampling as far as possible, the file of every patient tested in the year 1948 was examined consecutively and every file which met the criteria of containing the TAT and also the records of ten therapeutic interviews around the testing period was included. Since the year 1948 did not provide 50 cases, the files of the following year were examined in the same way until 50 such records were obtained. The 50 records used included 33 of males and 17 of females. It should also be noted that the specific cards used in the testing varied according to the judgment of the examiner; hence not all subjects were presented with each of the cards included in the study. Cards 1, 2, 6 BM, and 7 BM were used most frequently.

On the basis of a review of a sample of TAT stories and therapeutic interviews of 20 patients not used in the study, a check sheet was devised to include the 12 attitudes by which these patients characterized parental figures in the stories and in therapy sessions. The sheet provided a separate check list for each parent. Thus, it could be indicated that in the story in response to TAT Card 1, for example, a dominating attitude was attributed to the father, or that in response to 6 BM a solicitous attitude was attributed to the mother, etc. An identical check sheet was used to note the attitude mentioned during therapy sessions, and on this sheet the interviews (from 1 to 10) were indicated.

The attitudes included on the check sheets were as follows:

1. Comforting, giving, affectionate
2. Solicitous, concerned
3. Advice giving
4. Indifferent
5. Excessively attached
6. Over-giving
7. Dependent
8. Self-centered

9. Demanding (includes nagging; to get things for the parent)
10. Dominating, authoritative, wants to control, direct
11. Scolding, punishing
12. Rejecting, critical, obstructing, disapproving.

In the same manner a check list of images in which parental figures were depicted was also designed. These were:

1. Attractive, good, appetizing
2. Competent, effective
3. Incompetent, ineffective
4. Bad, evil, unattractive
5. Friction between parents.

Thus the check list included a total of 17 items.

The TAT stories and the ten therapeutic interviews were then analyzed by the two investigators together.[2] Every time that the presence of an attitude or an image in stories for each TAT card or in each therapy interview was noted, the item was marked on the pertinent check list in relation to the parent to whom it applied. If no attitude or image was discernible in a story on a TAT card or in an interview, this fact was indicated. The TAT stories and therapeutic interviews were analyzed independently, without knowledge of which stories were associated with which interviews.

Results

The data were analyzed to answer the three questions raised.

1. Do TAT stories serve as a basis for prediction as to whether or not patients who discuss parental attitudes and images in the test will discuss them also in therapy sessions? A chi-square analysis yielded no significant results in regard to:

Fathers' attitudes
Mothers' attitudes
Fathers' images
Mothers' images.

The results indicated clearly that the TAT furnished no basis for prediction in this situation.

2. Do the attitudes ascribed to each parent in the TAT resemble those expressed in the therapy sessions? Analysis[3] of the data gave the following results:

1. In relation to fathers' attitudes, only six records out of the 50 had any item in common in the TAT and therapy; none had more than one item. Of the six records, only three could be interpreted as challenging the null hypothesis at a level of confidence between 5 and 10 per cent.
2. In regard to mothers' attitudes, 15 of the 50 records had one item in common and one had two in common. Only the latter could be interpreted as challenging the null hypothesis at a level of confidence between 5 and 10 per cent.

Hence, the results indicated that the kind of attitudes as related to parents in TAT stories furnished no basis for prediction of those that were ascribed to them in therapy sessions.

3. Do the images of parents as noted in TAT stories correspond to those expressed in therapy? Analysis of the data in answer to this question gave the following results:

1. In relation to fathers' images, only five records out of the 50 had any item in common, only one of which was significant at the 10 per cent level.
2. Of mothers' images, only five had any item in common, none more than one, and none reached even a 10 per cent level of significance.

Thus, analysis of the data bearing on all three of our inquiries corroborated a null hypothesis of little or no association between material in the TAT and material in the therapeutic interviews.

It seemed possible that one explanation of this low degree of association might result from the fact that the attitudes and images itemized on the check list were not sufficiently exclusive to permit precise differentiation by the investigators. To explore this possibility, the items on the check lists were divided arbitrarily into two groups, one group containing "positive" or favorable attitudes and images, the other "negative" or unfavorable ones.

---

[2] A sample of ten cases involving 330 judgments was analyzed by the investigators separately in order to obtain a measure of reliability. Phi coefficients of reliability were .54 for TAT protocols and .84 for the therapy sessions. Both of these values are significant considerably beyond the .001 level.

[3] A combinatorial analysis employing the following formula, suggested by Dr. Joseph Gengerelli, was applied to the data:

$$\left[\frac{C^{M-(N_c+N_b)}_{N_a}}{C^{M}_{N_c+N_a}}\right]\cdot\left[\frac{C^{M-(N_c+N_a)}_{N_b}}{C^{M}_{N_c+N_b}}\right]\cdot\left[C^{M}_{N_c}\right]$$

where M = total number of possible choices, N_a = unique choices in TAT, N_b = unique choices in therapy, and N_c = choices in common.

This procedure, however, resulted in only a negligible increase in predictive value.

In the course of this dichotomized type of analysis, an interesting trend was noted. It was observed that some patients ascribed both positive and negative attitudes to the same parent. For example, a father would be described as "comforting and giving" at one point, and as "rejecting" or "dominating" at another. This "ambivalence," as we called it, occurred both in TAT stories and in therapeutic interviews in relation to both parents. However, in both the TAT and in therapy sessions the ambivalence expressed in regard to mother figures was significantly[4] greater than that in relation to fathers. This finding is consistent with many writings which have suggested that in our culture the relationship with the mother is the more complex and the more important in emotional illness.

Discussion

In interpreting the negative findings reported in this study, two considerations should be taken into account.

In the first place, it has already been mentioned that not all of the TAT cards selected for study were administered to each patient. The incompleteness of TAT data reduced the opportunity for producing material. Had additional material been available, it might have increased the value of the TAT in the predictions discussed.

In the second place, there is a possibility that TAT pictures evoke descriptions of parental attitudes which are not conscious to the patient at the beginning of therapy, but which become conscious to him in the course of therapy. If this possibility is a valid consideration, then little or no relationship between attitudes expressed in TAT stories and in the early stages of therapy should be expected. However, then we would predict that the attitudes expressed at some later point in successful insight therapy would correspond more closely to those in the TAT stories. A follow-up study investigating this possibility would be desirable.

Summary

In this study relationships between parental attitudes and images expressed by patients to TAT cards and in psychotherapy were investigated. Findings suggest no predictability from TAT to therapy as to whether or not parents were discussed. When they were discussed, no similarity was found between those attitudes and images given in TAT stories and in psychotherapy. Ambivalent attitudes toward both parents were present, but significantly more often toward mothers.

Received June 11, 1954.

References

1. Bell, J. E. *Projective techniques.* New York: Longmans, Green, 1948.
2. Bellak, L. The Thematic Apperception Test in clinical use. In L. E. Abt & L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1950. Pp. 185–229.
3. Butler, Pearl. Parent figures in TAT records of children in disparate family situations. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1948.
4. Combs, A. W. The use of personal experience in the Thematic Apperception Test story plots. *J. clin. Psychol.,* 1946, **4**, 358–363.
5. Harrison, R. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients. II. A quantitative validity study. III. Validation by the method of "blind analysis." *Charact. & Pers.,* 1940, **9**, 122–138.
6. Markmann, R. Predictions of manifest personality trends by a thematic analysis of three pictures of the Thematic Apperception Test. Unpublished thesis, Radcliffe College, 1943.
7. Mitchell, H. E. Social class and race as factors affecting the role of the family in Thematic Apperception Test stories of males. Unpublished doctor's dissertation, Univer. of Pennsylvania, 1950.
8. Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
9. Murray, H. A. *Thematic Apperception Test. Manual.* Cambridge, Mass.: Harvard Univer. Press, 1943.
10. Rotter, J. B. Thematic Apperception Tests: suggestions for administration and interpretation. *J. Pers.,* 1946, **15**, 70–92.
11. Singer, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *J. consult. Psychol.,* 1954, **8**, 99–104.
12. Tomkins, S. S. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.

---

[4] Critical ratio of difference between proportions in TAT stories, 2.4; in therapy sessions, 3.2.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Application of a Psychometric Scoring Procedure to a Group Modification of the Thematic Apperception Test (N-TAT)[1]

Leonard D. Eron, Florence Sultan, and Frank Auld, Jr.
Yale University

This paper reports an extension of a psychometric scoring procedure (3), which was worked out with the standard TAT method, to a group TAT used by the Navy. The Navy TAT (N-TAT) has been described by Briggs (1) in a previous report.

Subjects and Procedure

The subjects for this study were 100 sailors chosen systematically from two classes at the U. S. Navy Submarine School at New London, Connecticut. The N-TAT protocol of every fifth man in these classes was drawn from the files until a sample of 100 had been obtained. Because the method of selection appears to be unbiased, we can assume that the group of 100 is representative of all submarine school entrants.

The N-TAT protocol consists of 10 stories written in a booklet on each page of which is a different stimulus picture. Seven of the pictures are slight modifications of the standard third edition series (8) and three are original.[2] For the purposes of this study, after each page of each booklet had been assigned an identifying number, the booklets were taken apart so that all stories for any given picture could be analyzed at the same time.

Each story was analyzed for the following variables: emotional tone, outcome, themes, perceptual distortions, unusual formal characteristics, and level of interpretation. All the stories were analyzed for a single variable at a time and scoring on one variable was completed for 100 stories before proceeding to the scoring of the next variable. Once such an analysis had been made, it was possible to compile both apperceptive and thematic norms (cf. Rosenzweig, 10), since the same scheme can be used for each picture and the scores on one picture are comparable to the scores on each of the other pictures. The reliability of the scoring scheme was tested by having two judges (Eron and Sultan) independently score 20 stories for each picture (a total of 200 stories for all 10 pictures). Results of the reliability study are reported below in connection with the discussion of each variable.

Results and Discussion

Emotional tone. In the analysis of the emotional tone of the stories, a set of five-point scales empirically derived for the standard set of pictures was utilized (6). Since three of the pictures are original with the N-TAT, the specific scales could be used for only seven and the general rating scale for emotional tone was used for the others. A Pearson r of .89 between two raters on 200 stories

---

[1] This study was supported by the Office of Naval Research, through the American Council on Education, and by the Foundations Fund for Research in Psychiatry. The conclusions reported here do not necessarily reflect the official opinion of the U. S. Navy. Thanks are extended to Commander G. J. Duffner of the U. S. Naval Submarine Base for his cooperation and support.

[2] The actual plates were reproduced in the earlier article (1). They are identified for our purposes as follows: 1. "The Operation," 8 BM in the 3rd edition. 2. "10 Minute Break," 9 BM in the 3rd edition. 3. "Father-Son," 7 BM in the 3rd edition. 4. "Captain's Mast," original. 5. "Submarine in Storm," adapted from 19 in the 3rd edition. 6. "Morning After," 13 MF in the 3rd edition. 7. "Contemplation," adapted from 14 in the 3rd edition. 8. "Pickup," original. 9. "Persuasion," 4 in the 3rd edition. 10. "Seabag," original.

Bureau Edtnl. Psy. Research
Dated
No

Table 1

Percentage Frequency of Stories of Various Emotional-Tone Categories for Each of the TAT Pictures in the Navy Series

Picture no.	Emotional-tone rating					
	−2	−1	0	+1	+2	Reject
1	10(13.3)*	79(40.0)	5(9.3)	6(37.3)		
2	2(0.0)	24(26.7)	41(48.0)	33(24.0)	0.0(1.3)	
3	2(5.3)	49(58.7)	33(32.0)	14(4.0)		2(0.0)
4	1	92	2	5		
5	3(9.3)	61(33.3)	23(14.7)	3(33.3)	0.0(6.7)	10(2.7)
6	12(45.3)	70(49.3)	11(5.3)	6(0.0)		1(0.0)
7	8(9.3)	70(18.7)	10(45.3)	12(22.7)	0.0(4.0)	
8		61	32	6		1
9	4(4.0)	81(94.7)	7(1.3)	6(0.0)		2(0.0)
10		6	17	73	3	1

* Figures in parentheses refer to percentage frequency in original standardization group (3).

compares favorably with the reliability estimates obtained in other investigations using these scales with male and female college students ($r = .86$) (6) and adolescent girls ($r = .93$) (7). Table 1 presents the emotional-tone ratings for stories written by 100 sailors in response to each of the 10 N-TAT pictures.

For purposes of comparison, the percentage of ratings in each emotional-tone category in stories contributed in response to the corresponding pictures by 75 college student subjects of similar age and sex to whom the Murray cards were individually administered in the standard manner is included in parentheses. There are marked and statistically significant differences (beyond the .001 level of confidence) in the distribution of emotional-tone ratings between the two groups of subjects on three of the pictures. On picture 1 (The Operation) the college students contribute more +1 and fewer −1 stories. This no doubt occurs because the college group has more stories of aspiration toward a medical career and the sailor group has more stories about casualties at sea. For picture 5 (Submarine in Storm) there is the same kind of difference. Since this picture is a definite alteration of 19 in the standard series, designed to make the stimulus look like a submarine in distress, the sailors have more of this kind of story than of the usual story about the warmth and comfort of home in the midst of storms, which appears among the college students for that picture. The story of the comforts of home is, of course, a more positively toned interpretation than the one about a submarine in distress. On picture 6 (Morning After) the sailors contribute more −1, and the college students more −2 stories. That the college group gives sadder stories for this picture perhaps reflects the fact that in this group the theme of death of the wife is equally as prominent as the theme of remorse after intercourse, whereas among the sailors the latter is the modal theme. This may reflect a personality difference in the two groups. For picture 7 (Contemplation) the sailors have more −1 stories and the college group more 0 stories. This may be a function of the alteration in the picture—from a figure looking out of a window to a sailor leaning over a rail.

Table 2 presents thematic norms for emotional tone—the percentage of subjects who have a given number of stories of each emotional-tone rating in a series of 10 stories. There is no real basis for comparison with thematic norms previously compiled in other groups, since they have been based on all 20 pictures in the standard series. However, it is obvious that, just as in the individually ad-

Table 2

Percentage Frequency of Individuals Who Have a Given Number of Stories of Each Emotional-Tone Category in a Total of Ten Stories Given in Response to the Navy TAT

No. of stories	Emotional-tone rating							
	−2	−1	0	+1	+2	Reject	X	?
0	67		20	9	97	93	99	94
1	26	2	27	49	3	6	1	4
2	5		27	22		1		2
3	2	2	11	12				
4		12	11	5				
5		18	2	2				
6		31		1				
7		26	1					
8		6	1					
9		3						

ministered TAT, these pictures too elicit predominantly sad stories.

Outcome. A set of five-point scales empirically derived on the standard series of pictures in another investigation (3) was used in the analysis of outcomes. The inter-interpreter reliability derived from the independent ratings of 200 randomly selected stories by two judges was .81 (Pearson *r*), which compares favorably with reliabilities obtained in three other studies (3, 4, 7). It was necessary to add an additional category, *X*, to the scale to cover those stories in which there was no mention of any outcome. In the individual administration it is always possible to ask for an outcome when the subject fails to give one. This is, of course, impossible in a group administration. Thus only when the subject gives alternative outcomes or specifically writes that he cannot make one up is he scored with a *?* outcome in this procedure. Otherwise, if there is no definite outcome, he is scored *X*. Consequently, in this group of subjects the largest category for any story was *X*, and a comparison with the standardization group is not valid. However, the outcome ratings of all stories contributed by the Navy group for each of the pictures are provided in Table 3 as normative data to be consulted when the N-TAT or a similar kind of adaptation is used. Similarly, thematic norms (for a series of 10 pictures) are provided in Table 4.

Themes. In analyzing the themes, the second edition of the check list, the derivation of which has been described elsewhere (2), was used. Since there are three new pictures (Captain's Mast, Streetcorner Pickup, and Seabag) in this series, it was necessary to revise the check list further and add some new themes.[3] However, the number of changes necessary was not significant. Reliability of the "Navy" check list was good. Percentage agreement between two independent raters was calculated according to the following formula:

$$\frac{\text{(No. of agreements between Judge A and Judge B)} \times 2}{\text{Total judgments A + Total judgments B}}.$$

It was 89 per cent for a total of 200 stories (20 stories from each of 10 pictures). This compares well with what has been obtained in another study by different investigators (7).

The most frequent themes for each picture contributed by 10 per cent or more of the subjects in this group of 100 sailors or in the standardization group (for the comparable pictures only) are listed

Table 3

Percentage Frequency of Stories of Various Outcome Categories for Each of the TAT Pictures in the Navy Series

Picture no.	Outcome rating							
	−2	−1	0	+1	+2	Reject	*X*	*?*
1	10	6	1	21	2		49	11
2	1	18	26	13	1		37	4
3	2	10	8	19	1	1	54	6
4	1	42		19	2		26	10
5	3	5	7	15	2	8	57	3
6	9	23	9	7	2	1	35	14
7	5	8	8	21			53	5
8		21	11	13		1	41	13
9	3	12	5	19		2	54	6
10		4	6	44	4	1	33	8

Table 4

Percentage Frequency of Individuals Who Have a Given Number of Stories of Each Outcome Category in a Total of Ten Stories Given in Response to the Navy TAT

No. of stories	Outcome rating							
	−2	−1	0	+1	+2	Reject	*X*	*?*
0	74	26	49	21	92	91	15	47
1	19	30	31	33	5	5	9	32
2	6	24	14	17		3	10	15
3	1	12	5	12	3	1	6	6
4		6		7			11	
5		1		5			8	
6			1	5			13	
7		1					12	
8							8	
9							5	
10							3	

---

[3] The Navy check list of themes, defining each theme and noting the changes, deletions, and additions to the second edition, has been deposited with the American Documentation Institute. Order Document No. 4421 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

in Table 5. The chief differences between the two groups were alluded to in the discussion of emotional tone. They appear to be largely a function of the physical differences between the original Murray pictures and the adaptations used in this study, as well as of the different life situations and predominant needs of the subjects. Because of the two variables—the stimuli and the subjects—which are confounded, it is impossible to partial out the exact cause of the differences in results. However, those who use this or similar adaptations of the TAT with similar subjects should be aware of these differences when making interpretations. Thematic norms appear in Table 6.

Levels of interpretation.[4] It was necessary to develop a new scale for levels of interpretation in this study because of the difference in administration from the standard method. Stories were designated on an eight-point scale going from complete rejection of the picture to complete narrative (including a

---

[4] This category does not refer to a variable which carries a similar name in previous publications (2, 3, 4). The latter variable is now called "unusual formal characteristics."

Table 5

Most Frequent Themes in Stories Given in Response to Each of the TAT Pictures in the Navy Series*

Picture	Theme		Percentage frequency	
			Navy	College
1 (The Operation)	IA3i	Death or illness of peer	31	2.7†
	IB12	Death or illness of central character	25	14.7
	IC6	War	21	29.3
	IB29	Death, illness—unspecified character	18	
	IB14	Reminiscence, sad	17	6.7
	IB7	Guilt	15	
	IC5	Aggression from—impersonal	12	8.0
	IA3g	Concern—peer	11	
	IB1	Aspiration	6	53.3†
	IA1a	Death, illness of parent	7	18.7†
2 (10 Minute Break)	IB4	Retirement	47	68.0†
	IB21	Exhaustion	45	14.7†
	IIB3	Reminiscence, happy	31	4.0†
	IC7	Escape from perilous environment	13	
	IC6	War	10	9.3
	IB19	Vacillation		13.3
	IIB2	Tranquillity		16.0
3 (Father-Son)	IAib	Succorance from parent	62	74.7†
	IA11	Disappointment in parent	14	
	IA1a	Pressure from parents	6	28.0†
	IB11	Occupational concern	4	25.3†
4 (Captain's Mast)	IA1a	Pressure from parent	92	
	IC7	Escape from perilous environment	35	
	IB13	Retribution	31	
	IC4	Aggression toward environment	25	
	IB7	Guilt	17	
5 (Submarine in Storm)	IC5	Aggression from—impersonal	42	61.3†
	IC6	War	26	5.3†
	IIA1f	Contentment at home		26.7
	IB9	Fear		10.7
6 (Morning After)	IA2h	Illicit sex	64	42.7†
	IB7	Guilt	35	54.7†
	IB8	Drunkenness	17	13.3
	IA2j	Death, illness of partner, female	15	42.7†
	IA2v	Venereal disease, partner	14	
	IA2e	Aggression toward partner		17.3
	IB13	Retribution		16.0

Table 5—*Continued*

Picture		Theme	Percentage frequency	
			Navy	College
7 (Contemplation)	IIB3	Reminiscence, happy	27	12.0†
	IB28	Homesickness	22	
	IB17	Loneliness	11	5.3
	IA2s	Unrequited male partner	10	
	IB5	Suicide	10	2.7
	IB3	Curiosity		26.7
	IB1	Aspiration		22.7
	IB4	Behavior disorder		14.7
	IB10	Rumination		12.0
	IIC1	Favorable environment	5	10.7
	IIB2	Tranquillity		10.7
8 (Pickup)	IIB7	Ordinary activity	34	
	IA2pp	Decision, female partner	23	
	IA2n	Competition, partner	21	
	IA2q	Pursuit, partner	17	
	IA1a	Pressure, parent	15	
	IA11	Disappointment in, parent	15	
	IC2	Legal restriction	11	
	IA1b	Succorance, parent	10	
9 (Persuasion)	IA2r	Seduction	24	10.7†
	IA2c	Succorance, partner	22	22.7
	IA2a	Pressure, partner	19	50.7†
	IA2xx	Ulterior motive	14	
	IA2f	Departure from partner, female	13	16.8
	IB12	Illness or death central character	13	
	IA2h	Illicit sex	9	12.0
	IA2m	Competition		12.0
	IB6	Moral struggle		12.0
10 (Seabag)	IIA1d	Reunion, parent	51	
	IIB3	Reminiscence, happy	27	
	IB1	Aspiration	13	
	IIB7	Ordinary activity	13	
	IIB5	Occupational satisfaction	12	

* Only those themes contributed by 10 per cent or more of the subjects in either group are included in this table.

† Themes for which the difference between groups is significant beyond the .05 level of confidence.

conclusion and interpretation of feelings and thoughts of the characters). The scale appears in Table 7. The percentage agreement between two raters for 200 stories (20 stories for each of 10 pictures) was 85. The incidence of the various levels for each story appears in Table 8. From this table and Table 9, which presents thematic norms, it can be seen that, although certain of the pictures tend to elicit less well-developed stories, the general level of most of the responses indicates a good deal of involvement on the part of the majority of subjects, even in this group method of administration.

Perceptual distortions. From the number of perceptual distortions listed in Table 10 it can be seen that perceptual distortions are not nearly so frequent among the group of sailors as they were among the original standardization group (3). The lack of distortions is probably due to group administration, because the subjects can study the pictures more carefully and need not write down their first impressions. Also, since a new set of pictures was drawn up for this series, it may be that these pictures differ in ambiguity from the original set. One outstanding difference in results between this group and previous groups is in the number of rejections. In the previous studies such behavior was rare even among subjects who were psychiatrically ill (2, 3); in the Navy group, 10 per cent of the

Table 6

Percentage Frequency of Individuals Who Have a Given Number of Specific Themes in a Total of Ten Stories*

No. of theme	IA1a†	IA1b	IA1g	IA1i	IA1d	IA2g	IA2h	IA2k	IA2q	IA2r	IA2u	IA2v‡	IA2xx‡	IA3b	IA3c	IA3i	IA3j	IA3l
0	9	29	96	91	74	89	32	92	82	68	92	81	83	93	89	67	93	94
1	62	54	3	8	24	9	55	7	13	30	7	17	15	6	10	31	6	5
2	25	15	1	1	2	2	10	1	4	2	1	1	2	1	1	2	1	1
3	3	2					2		1									
4	1						1					1						

No. of theme	IB1	IB2	IB3	IB4	IB5	IB6	IB7	IB8	IB9	IB10	IB11	IB12	IB13	IB14	IB16	IB17	IB21‡	IB34‡
0	72	89	91	85	89	95	47	61	72	95	85	59	65	74	89	88	53	85
1	22	9	5	14	10	4	33	32	25	4	14	36	33	18	10	11	43	14
2	5	2	4	1	1	1	18	4	3	1	1	5	2	6		1	4	1
3	1						2	3						2	1			
4																		

No. of theme	IC2‡	IC3	IC4‡	IC5	IC6	IC7‡	IC8	IIA1d	IIA3d	IIA2e	IIB3	IIB5	IIB7	IIC1
0	81	94	70	45	57	52	93	51	91	92	40	86	50	89
1	17	5	28	50	27	39	6	44	8	7	40	13	39	9
2	2	1	2	5	7	9	1	5	1	1	15	1	10	2
3					5						4		1	
4					4						1			

* Only those themes contributed at least twice by any subject are included in this table.

† Themes given here by code number have been defined in Table 1 in a previous article published in this journal (1953, 17 [4], 269–282). The themes indicated by ‡ are original with the N-TAT and thus do not appear in that list. They are: IA2v, venereal disease; IA2xx, ulterior motive; IB21, exhaustion; IB34, grief; IC2, legal restriction; IC4, aggression toward environment; IC7, escape from perilous environment. Definitions of these themes appear in the Navy check list deposited with the American Documentation Institute (see footnote 3).

subjects rejected at least one picture. The increased rejection rate is probably a function of the group administration method. It is easier to insure maximum cooperation throughout an entire test series when the experimenter is present to prompt and encourage the subject throughout as in the individual administration. With the present method it is difficult to assess the exact cause of this large number of rejections.

Table 7

Level of Interpretation

A	Rejection
B	Description of physical appearance of objects and/or persons
C	Description of action
D	Description of feelings, attitudes, and/or thoughts of characters
E	Incomplete narrative—omits either antecedents, present action, or conclusion
F	Complete narrative—includes antecedents, present action, and conclusion
G	Incomplete narrative with interpretation of feelings, attitudes, and/or thoughts
H	Complete narrative with interpretation of feelings, attitudes, and/or thoughts
W	Not really a story, reflection on life, words, advice, remonstrance

Application of the Method

The practical utility and validity of this type of psychometric analysis of the N-TAT

Table 8

Percentage Frequency of Stories of Various Level-of-Interpretation Categories for Each of the TAT Pictures in the Navy Series

Picture no.	Level rating								
	A	B	C	D	E	F	G	H	W
1			6	8	2	11	33	35	5
2			7	5	12	27	12	32	5
3	1		10	2	27	16	12	29	3
4			4	3	12	30	7	40	4
5	8	15	24		10	27	5	6	5
6	1		7	6	7	6	16	50	7
7			5	18	3	5	24	39	6
8	1		8		18	26	12	30	5
9	2		13	2	9	11	28	31	4
10	1		13	8	2	18	15	35	8

Table 9

Percentage Frequency of Individuals Who Have a Given Number of Stories of Each Level-of-Interpretation Category in a Total of Ten Stories Given in Response to the Navy TAT

No. of stories	Level rating								
	A	B	C	D	E	F	G	H	W
0	91	85	60	61	41	28	33	12	82
1	5	15	16	28	30	26	23	21	9
2	3		12	9	16	17	20	14	3
3	1		4	2	11	15	7	14	2
4			2		1	7	8	5	1
5			1		1	3	7	11	
6			4			2	2	8	
7						2		9	
8			1					3	1
9								2	1
10								1	1

were demonstrated in a study the results of which are fully reported elsewhere (5). A comparison was made of TAT stories told by 22 sailors before they were sealed up in a submarine and the stories told by them later, while they were still enclosed. It was possible to demonstrate, on the basis of the protocols, in what ways the 22 men who volunteered for the experiment ("Operation Hideout") differed from the usual run of submariners, and also to enumerate the differences which appeared in their fantasy productions after the experimental treatment. These changes in fantasy coincided with differences observed in physical, physiological, and psychomotor functions.

Table 10

Percentage Frequency of Perceptual Distortions for Each of the TAT Pictures in the Navy Series

Picture no.	Distortion	Frequency
1	Patient as female	13
	Picture within picture	1
2	Responds to only one character	2
3	Self-image	1
4		
5	Stove or gun cover	1
	Serpent	1
	Fish	3
	Rabbit	1
	Locomotive—cat face	1
	Dragon	1
6		
7		
8		
9	Female seen as mother	1
10	Gun with arms and legs	1

Summary and Conclusions

A psychometric method of analysis of TAT stories derived from protocols obtained in the standard manner was extended to a group adaptation of the TAT which is used by the U. S. Navy. The test records of 100 systematically selected submariners were analyzed for emotional tone, outcome, themes, level of interpretation, and perceptual distortions. The reliability of the scales and check lists used was demonstrated to be of sufficiently high order to warrant their use with this type of administration and this type of subject. Both apperceptive and thematic norms are provided for the convenience of those using this adaptation. This psychometric procedure was tried and found to be useful and valid in a research which is reported in a separate monograph.

Received June 16, 1954.

References

1. Briggs, D. L. A modification of the Thematic Apperception Test for naval enlisted personnel (N-TAT). *J. Psychol.*, 1954, **37**, 233–241.
2. Eron, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. consult. Psychol.*, 1948, **12**, 387–395.
3. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
4. Eron, L. D. Responses of women to the Thematic Apperception Test. *J. consult. Psychol.*, 1953, **17**, 269–282.
5. Eron, L. D., Auld, F., Jr., Lesser, G., Laffal, J., & Sultan, Florence. A study of the TAT stories and sentence completions of subjects in "Operation Hideout." *Bur. Med. Surg., Navy Dept., M.R.L. Rep.*, 1954, No. 243.
6. Eron, L. D., Terry, Dorothy, & Callahan, R. The use of rating scales for emotional tone of TAT stories. *J. consult. Psychol.*, 1950, **14**, 473–478.
7. Liccione, J. V. The changing family relationships of adolescent girls. Unpublished doctor's dissertation, Syracuse Univer., 1954.
8. Murray, H. A. *Manual for the Thematic Apperception Test*. Cambridge, Mass.: Harvard Univer. Press, 1943.
9. Ritter, Anne M., & Eron, L. D. The use of the Thematic Apperception Test to differentiate normal from abnormal groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 147–158.
10. Rosenzweig, S. Apperceptive norms for the Thematic Apperception Test. I. The problem of norms in projective methods. *J. Pers.*, 1949, **17**, 475–482.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Effects of Sexual Stimulation Under Varying Conditions on TAT Sexual Responsiveness[1]

Paul H. Mussen and Alvin Scodel
The Ohio State University

The present study is concerned with the effects of sexual stimulation in different situational contexts on subsequent TAT sexual responsiveness. Our hypothesis states that these responses will be more inhibited following sex arousal in the presence of a stern father figure than following such arousal in the presence of a young, permissive individual.

A procedure almost identical with the one described by Clark (1) was used. The subjects were two groups of male students in an introductory psychology course. Both groups took the TAT after having been exposed to sex-arousing stimulation. In the Formal group (33 subjects), a series of eight photographic slides of attractive nude females was presented by a formal, professorial, and somewhat stern man in his sixties. Subjects were told that the research was related to Sheldon's body-type theories. They were instructed to rate the attractiveness of each of the eight nudes and to indicate the criteria for their judgments.

Using the same instructions, a young-looking, informal, permissive graduate student presented the slides to the Informal group (38 subjects).

In both groups, the first examiner left the room immediately after the ratings were completed. At this point, another experimenter stepped forward, explained that he was attempting to standardize a test of creative imagination, and then administered eight TAT cards from the Murray series (7BM, 12BG, 3BM, 14, 20, 6GF, 6BM, 13MF) to the group.

The manifest and implied sexual content of each TAT story was scored according to a seven-point scale, ranging from 0 to 6, depending on the intensity of sexual activity. The subject's total sexual response score was the sum of the scores on the eight stories.

The mean sexual response score of the Informal group was 9.55, while the mean for the Formal group was 7.12 (*SD*'s, 5.18 and 4.26, respectively). The difference between means yielded a t value of 2.13, significant between the 2 and 5 per cent levels of confidence.

Thus, it would seem that arousal of socially disapproved needs in the presence of an authority figure results in greater inhibition of subsequent expression of that need than arousal of the same need in a permissive situation.

Brief Report
Received November 11, 1954.

Reference

1. Clark, R. A. Projective measurement of sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391–399.

---

[1] An extended report of this study may be obtained without charge from Paul H. Mussen or Alvin Scodel, Ohio State University, Columbus 10, Ohio, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4439 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Diagnostic Significance of Usual and Unusual TAT Stories

Jerome S. Miller and Alvin Scodel
The Ohio State University

In spite of numerous efforts, direct and frontal attempts to validate the TAT have not been convincingly successful. Contributing to this difficulty are such factors as variations in administration and scoring procedures employed by users of the TAT and differences in interpretation as a function of the clinician's orientation and purpose. But, perhaps most important, the major problem in validation of the TAT lies in the elusiveness of adequate criterion data, combined with a lack of clinical or theoretical agreement as to just what it is that is to be validated.

Despite these obstacles to its demonstrable acceptability, other approaches are available for determining the TAT's efficiency for eliciting meaningful interpretive data. For example, construct validity, or the principles upon which such an instrument is based, often can be put to test. Demonstrating the plausibility of underlying assumptions is another way which also can lend to the credibility of inferences from such an instrument.

Lindzey has recently published an article exploring the latter of these two alternatives (4). He examined a number of pertinent assumptions underlying the TAT and evaluated them in the light of empirical evidence available in the literature. Our present study was designed to extend Lindzey's alternative method by isolating one underlying assumption and testing it within a TAT situation.

The specific assumption to be tested holds that stimulus-bound responses are not as diagnostically significant as responses which are less constrained by the stimulus material. Lindzey stated that this assumption—" . . . Themes or Story Elements That Appear to Arise Directly out of the Stimulus Material [Are] Less Significant than Those That Are Not Directly Determined by the Stimulus Material . . . "—is consistent with data from the investigations of "new look" perceptionists (4, pp. 13 f.).

Stated in another way, the assumption we studied refers to the significance of the unusualness of a response, i.e., the degree to which an idiosyncratic or atypical fantasy production is diagnostically useful. One recent study states in the introduction that, "Despite the scarcity of normative data, various clinical investigators have stated that the presence of certain characteristics or 'deviant' content in a TAT record provides a basis for differential diagnosis" (8, p. 147). The same article concludes " . . . in the interpretation of a given protocol it is felt that this study offers further evidence that fantasy material is important chiefly when it deviates from the expected response" (p. 157).

Rotter states that "Occasionally the subject tells a story which has little or no significance in terms of his basic problems. Faced with a stimulus which he does not react to emotionally, he gives some common interpretation based on relatively superficial experiences with such situations" (9, pp. 83–84). Murray makes a somewhat similar point when he emphasizes the interpreter's task of " . . . noting everything that is unusual, uncommon or unique; or common but unusually high or low in intensity or frequency (5, pp. 7–8). Or as put by Rapaport *et al.*, "The standard cliché is always approximated but never reached; it is an abstraction from which deviations may be measured. The greater the deviation, the more amenable is the story to diagnostic interpretation" (7, p.

421). In general then, many psychologists interpret TAT themes on the basis of variation from implicit or explicit norms of typical themes.

Method

Normative data for the TAT have been notably lacking. Therefore, there are no available standards by which to measure the degree to which a given response is determined by the stimulus material. Moreover, there are no empirical norms representing the common or usual response elements for each stimulus. An exception to this deficiency exists in the introductory set of card-by-card norms recently published by Eron (1). These norms encompass a rating scale for emotional tone of TAT stories; a frequency count of descriptive theme categories and interpretation levels; outcome ratings of stories; frequency of direction of shift in emotional tone from the body of the story to the outcome; identification of characters and objects; and the extent to which subjects single out some unusual or minute aspect of the picture for elaboration in their stories. The data were collected from a total sample of 150 male veterans of World War II.

An earlier attempt by the present writers to test the hypothesis of the diagnostic significance of thematic unusualness, using Eron's normative data, was unsuccessful. The plan, in brief, was to obtain instances where two individuals had each given a story subsumed under a single category except that each instance was in response to different stimulus cards. One production was considered atypical in terms of the normative frequency data and the other story was considered as typical. However, several attempts to obtain reliability between judges in labeling the thematic categories of stories resulted in only 50 per cent agreement and the plan was therefore rejected.

Selection of the story population. For the procedure used in the present study 35 sets of TAT stories were obtained. Each set contained three stories given to each of five TAT stimulus cards (1, 6BM, 7BM, 12M, and 16). Each subject's successive stories to any one card were given in response to instructions to "make up another story to this card, but make this story as different as possible from the story you made up before for the same picture." Subjects were male adults and the 35 represent all of such specially tested men who were in psychotherapy for two or more treatment sessions at a mental hygiene clinic. At the time of this study all 35 had discontinued treatment and were "closed" cases.

Table 1

Distribution of Sums of Squares of Two Judges Ranking Differences for a Five-Point Scale ($N = 35$)

Sums of differences squared	Frequencies		
	Theoretical	Observed	D^2/ Theoretical
0–10	7.87	29.0	56.73
12–16	5.84	2.0	2.53
18–22	7.59	4.0	1.70
24–28	5.84	0.0	5.84
30–40	7.87	0.0	7.87

Note.—$\chi^2 = 74.67$ ($df = 4$), $p < .0001$, $C = .72$.

These stories were presented to two experienced clinicians who independently ranked the stories on a five-point scale, the most unusual story in a protocol being ranked "1" and the least unusual (or most usual) story "5." Only the first stories given to each stimulus card were used. The two judges each employed their own private, implicit norms for this task.

Reliability of selection was measured by using Olds' form for the distribution of sums of squares of rank differences (6). A chi square was computed between this expected distribution and the distribution of D^2 between the two judges' rankings of the stories. This yielded a chi square of 74.67 with $df = 4$ and $p < .0001$ as shown in Table 1. Where the two selection judges were not in exact agreement at the extremes of the scale, a third experienced judge arbitrated the disagreements. In this way 35 usual and 35 unusual TAT stories were obtained.

Procedure. The 70 stories were typed on individual cards without further identifying data and presented to six new judges. The judges attempted "blind" matching to typescripts of the first two psychotherapy sessions of each of the 35 subjects. The cards and the psychotherapy records were presented to judges in groups. Each group consisted of either three or four sets of therapy records and the two stories—the usual and the unusual—corresponding to each of those records. The task was to match each of the stories with one of the records. The judges were unaware that there were two kinds of stories. They were aware, however, that the stories had been obtained originally so that it was possible for two different stories from the same stimulus card to have been given by one subject. The number of stories matched with any case record ranged from none to five.

Another precaution was to equate roughly for length all stories within any particular group. This was done to reduce the possibility of allowing productivity or verbality to serve as a basis for matching.

Results

The number of correct and incorrect matchings for the usual and unusual TAT stories are shown in Table 2.[1] Chi squares were computed for each judge separately. The over-all significance of combined results was also computed using the chi-square model of Jones and Fiske (3) on the assumption that each judge's results are unrelated to the probability associated with each of the other judges' results. This combined result provides a chi square of 12.46 which, with $df = 12$, gives a p value of .41 approximately (Table 2).

Table 2

Results of Judges in Matching Usual and Unusual TAT Stories with Psychotherapy Records

	Judge					
	A	B	C	D	E	F
Number of usual						
Matched	11	13	17	19	11	17
Unmatched	24	22	18	16	23	17
Number of unusual						
Matched	14	15	17	15	17	21
Unmatched	21	20	18	20	17	13
χ^2 ($df = 1$)	0.58	0.24	0.00	2.04	2.18	0.96
p	.46	.67	1.00	.15	.13	.33

Note.—Combined $\chi^2 = 12.46$ ($df = 12$), $p > .40$ (3).

Similarly, chi squares were computed after eliminating instances where a judge had correctly assigned both the usual and unusual stories (Table 3). This was done to determine whether or not some of the lack of difference might have been due to matching mediated by extraneous cues. A usual story, in some cases, might have been paired with its corresponding unusual story through similarity of language factors in the stories, administration techniques, the examiner's individual fashion of recording, etc., thus supervening direct matching to a therapy record. However, even with these cases deleted, the data fail to show a marked superiority for the matching of unusual stories.

Table 3

Results of Judges in Matching Usual and Unusual TAT Stories with Psychotherapy Records: Paired Matchings Eliminated

	Judge					
	A	B	C	D	E	F
Number of usual						
Matched	5	8	9	11	4	8
Unmatched	24	22	18	16	23	17
Number of unusual						
Matched	8	10	9	7	10	12
Unmatched	21	20	18	20	17	13
χ^2 ($df = 1$)	0.90	0.32	0.00	1.32	3.48	1.34
p	.35	.63	1.00	.26	.06	.26

Note.—Combined $\chi^2 = 14.04$ ($df = 12$), $p > .30$ (3).

To determine whether the frequency of correct matchings was high or low over all, a chi square was computed for usual and for unusual stories each versus chance expectation (Table 4). Chance frequency was 11 for each of the two kinds of stories for each judge. Three judges (C, E, F) matched unusual stories at better than the .05 level and three

Table 4

Judges' Matchings of Usual and Unusual TAT Stories: Each Versus Chance

	Judge					
Chi square	A	B	C	D	E	F
Usual vs. chance	0.00	.53	4.77*	8.49**	0.00	4.86*
Unusual vs. chance	1.20	2.12	4.77*	2.12	4.84*	13.44**

* Statistically significant at the 5% level.
** Statistically significant at the 1% level.

---

[1] Judges E and F matched against only 34 subjects since one individual resumed treatment and his data became unavailable. The chance probability of correct matching remained the same, however, for each judge.

(C, D, F) matched the usual stories at better than the .05 level. With the use of Wilkinson's table (10) the combined results show that both the usual and unusual stories were correctly matched more frequently than by chance at about the .002 level. Both usual and unusual stories then were apparently matched with better than chance accuracy, with neither type of story showing significant superiority.

Discussion

Essentially, our interest lay in a demonstration of a "back-door" approach to the problem of validating the TAT and similar fantasy-production techniques. Testing an underlying assumption is not, by any means, either a crucial test of the TAT's validity or a decisive illustration of such an alternative, back-door course. The surprising feature of the study is the lack of confirmation of an assumption generally accorded high face validity and wide subscription by clinicians.

Because of the questions provoked by the unexpected results, the data were further treated to discount the possibility of matching mediated by extraneous cues. To do this, those cases were excluded in which both the usual and the unusual stories were correctly identified with their common psychotherapy criterion. For the cases which remained, a statistically significant difference between the two kinds of stories still did not appear. In addition, as noted before, the number of successful matchings for each of the two kinds of stories exceeded chance expectations.

There are some speculative explanations of the accurate identification of usual stories with their corresponding psychotherapy records. One reason may be that stories labeled usual are unremarkable only in relation to other stories from the same protocol. Thus, our two populations of stories, differentiated by their relative ranks, actually may both have been loaded in the direction of atypicality.

However, the two selection judges showed high agreement in their independent rankings, and we must conclude that there were, in general, discernible differences between adjacent stories throughout the five ranks—not merely between the first and fifth ranks.

Secondly, no TAT card yields but one cliché response. For example, Eron found that stories to Card 1 fitted thematic categories of "aspiration," "parental pressure," and "belongingness" in 46.7, 41.3, and 25.3 per cent of his cases respectively (1). A judge or clinician might then derive a meaningful differentiating cue by noting which of several stereotyped responses is chosen by the subject. Also, even a hackneyed story can be suggestive by virtue of the intensity of some fairly common theme. Suppose, for example, that a subject has intense dependency needs. Then on Card 6BM, for example, we might expect that subject to depart only minimally, if at all, from a theme of dependency so commonly given in response to this card. Yet we might be able to differentiate this subject from another on the basis of how intensely his story reflects his need structure in this particular area. Thus, the cliché is also characteristic of a subject and may be revealing of him in a broad context. The presence of repeated stereotyped stories in a protocol may be indicative of a subject's method of handling his problem areas.[2]

In closing this explorative discussion, a final question is raised which concerns a difference between the task imposed upon the judges of this study and the procedure generally followed in a clinical interpretation. In the latter situation there is some pressure upon the worker to identify the cues and to integrate and verbalize hypotheses and diagnostic descriptions. None of these requirements existed for the blind matching procedure. It is conceivable that judges may have derived unverbalizable or subliminal cues from a story and matched them with similar unverbalizable or subliminal cues inferred from the psychotherapy typescripts. It is not known what effect this freedom from explicitness has upon the data. It is possible in a clinical setting, however, that the notion

[2] Holt remarks on this interpretative value of usual stories in his statement that "It may be said further that frequently a great deal more can be interpreted from a set of 'cliché stories,' than is frequently claimed. It is more difficult; it requires more experience and more sensitivity to subtle ways in which a stereotyped story differs from dead average; but . . . it is always possible to make some progress" (2, pp. 201–202).

of unusual stories having diagnostic advantage stems more from ease of interpretation than from demonstrably superior predictiveness.

Summary

In this study we attempted to test an assumption generally held to have significance in the interpretation of TAT stories and similar fantasy productions. The hypothesis tested was that judges using a blind matching technique will more often ascribe unusual TAT stories to corresponding psychotherapy records of subjects than they will correctly ascribe usual stories.

A usual story and an unusual story were reliably obtained from each of 35 psychotherapy clients. The results of the judges' matching attempts, however, did not confirm the hypothesis at a statistically significant level, although each type of story was matched significantly better than chance.

These results were briefly discussed and evaluated and it was tentatively concluded that the stereotyped, or cliché, TAT story is not without meaning characteristic of its producer and should not be overlooked as a resource for the clinician's understanding of a subject.

Received July 21, 1954.

References

1. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
2. Holt, R. R. The Thematic Apperception Test. In H. H. Anderson & G. L. Anderson (Eds.), *An introduction to projective techniques.* New York: Prentice-Hall, 1951. Pp. 181–229.
3. Jones, L. V., & Fiske, D. W. Methods for testing the significance of combined results. *Psychol. Bull.*, 1953, **50**, 375–382.
4. Lindzey, G. Thematic Apperception Test: interpretive assumptions and related empirical evidence. *Psychol. Bull.*, 1952, **49**, 1–25.
5. Murray, H. A., *et al. Thematic Apperception Test. Manual.* Cambridge, Mass.: Harvard Univer. Press, 1943.
6. Olds, E. G. Distribution of sums of squares of rank differences for small numbers of individuals. *Ann. math. Statist.*, 1938, **9**, 133–148.
7. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Vol. 2. Chicago: Year Book Publishers, 1946.
8. Ritter, A. M., & Eron, L. D. The use of the Thematic Apperception Test to differentiate normal from abnormal groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 147–158.
9. Rotter, J. B. Thematic Apperception Test: suggestions for administration and interpretation. *J. Pers.*, 1946, **15**, 70–92.
10. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156–158.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Diagnostic Discrimination of the "Basic Rorschach Score"

Ben C. Finney
VA Hospital, Palo Alto [1]

The diagnostic adequacy of the Basic Rorschach Score (1) was tested by comparing the mean scores of groups with different diagnoses which were collected and scored by McReynolds (2, 3). The protocols were scored according to the "Diagnostic Sign List" (2, p. 80). In order to compare the results with those reported by Buhler *et al.* (1, Table III, p. 12), the 1945 and 1949 samples of comparable diagnostic groups of Buhler were combined on the basis of the statistics reported, and the neurotic divisions were combined into one single neurotic group. The statistics for Buhler's and McReynolds' samples are reported in Table 1.

In McReynolds' samples, there were no significant differences between the Neurotic vs. Normal or between the Paranoid vs. Mixed Schizophrenic groups. There were significant differences, at better than the .01 level, between the Normal vs. Paretic, Normal vs. Combined Schizophrenic, and Neurotic vs. Paretic groups. The difference between the Neurotic and the Combined Schizophrenic groups was significant at the .05 level.

When McReynolds' and Buhler's groups were compared, there were no significant differences between the Schizophrenic groups but Buhler's Normal group was significantly higher than McReynolds' Normals and Buhler's Neurotics were significantly higher than either McReynolds' Normals or Neurotics.

Table 1

Basic Rorschach Scores

Diagnostic group	*N*	Mean	*SD*
Normals (McR)	111	− 7.12	12.74
Neurotics, Mixed (McR)	22	−10.00	13.00
Schiz. Par. (McR)	31	−15.00	14.13
Schiz. Mixed (McR)	33	−18.76	10.14
Schiz. Combined (McR)	64	−16.94	12.38
Paretics (McR)	24	−26.42	9.13
Normal (Buh)	89	18.90	6.03
Neurotics, Comb. (Buh)	176	8.90	7.20
Schiz. Par. (Buh)	29	− 9.64	13.07
Schiz. Mixed (Buh)	63	−19.42	11.75
Schiz. Comb. (Buh)	92	−16.89	12.28

These results suggest that despite similar diagnoses, McReynolds' and Buhler's samples were not from a similar population and that within McReynolds' population, the Basic Rorschach Score was not adequate for use in diagnosing individual patients.

Brief Report
Received November 29, 1954.

References

1. Buhler, Charlotte, Lefever, D. W., Kallstedt, F. E., & Peak, H. M. *Development of the Basic Rorschach Score: supplementary monograph.* Los Angeles, Calif.: Rorschach Standardization Study. No. 1, 1948.
2. Cass, W. A., & McReynolds, P. A contribution to Rorschach norms. *J. consult. Psychol.*, 1951, **15**, 178–184.
3. McReynolds, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, **46**, 131–141.

---

[1] An extended report of this study may be obtained without charge from Ben C. Finney, VA Hospital, Palo Alto, California, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4481 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Psychotherapeutic Technique, Professional Affiliation, and Experience Level[1]

Hans H. Strupp[2]
Washington, D. C.

To the extent that psychotherapy aspires to scientific status, it must be based upon rational procedures that are capable of precise formulation and communication. A corollary of this proposition is that psychotherapeutic operations are not teachable unless the requirement of communicability is met. To this end, psychotherapy must strive to make its working assumptions, hypotheses, and operations increasingly explicit and objective.

Previous research in this field has dealt primarily with changes in the client as a result of therapy, diagnostic studies of the person seeking help, and the development of process measures. Few investigations have concerned themselves with the person of the therapist, and the character of his contribution to the dyadic interpersonal situation known as psychotherapy.[3] Yet it is almost axiomatic—and has been recognized in psychoanalytic writings since the time of Freud—that the therapist's personality, attitudes, past interpersonal experiences, and emotional blind spots are among the prime determiners of his therapeutic operations. It is believed that a program of research focusing on the therapist's contribution to the treatment situation is potentially productive for advancing our knowledge of psychotherapy.

As a first step, it is proposed to conduct a series of exploratory studies directed at the therapist's verbal techniques. This research attempts to answer the questions: (*a*) What techniques are used by psychotherapists of two theoretical orientations (Rogerian and psychoanalytic)? (*b*) Are there systematic differences in technique attributable to professional affiliation and level of experience? (*c*) Are there similar differences with respect to the therapeutic problem to which the therapist is addressing himself at the moment? (*d*) What is the effect of the therapist's personal analysis upon his verbal operations? The first problem has been dealt with in the first article of this series (6); the second will occupy our attention in the present paper; the remaining ones will be the subject of subsequent publications.

Method

It was considered advisable to avoid the complexities of the actual therapeutic situation, and to study therapists' verbal behavior through the use of a preselected sample of patient communications. Thus, a series of 27 short paragraphs of patient statements was culled from published therapeutic interviews, and typed on individual cards. A bare minimum of background information preceded each statement. Included in the series, which was considered a fair cross section of verbalizations heard from neurotic patients in early interviews, were a variety of complaints,

---

[1] This article is the second in a series of investigations concerned with the objective study of psychotherapists' verbal operations, and represents one phase of the author's doctoral dissertation submitted to the Graduate Council of the George Washington University (5). Grateful acknowledgment is made to Professor Curtis E. Tuthill, Dr. E. Lakin Phillips, and Leon Salzman, M.D., for valuable suggestions and criticisms.

[2] While carrying out this work, the author held positions successively with the Department of the Air Force and the Department of the Army. Neither organization has any connection with this research.

[3] In Mowrer's (3) recent compendium only one chapter (Fiedler's) specifically addresses itself to this problem.

Table 1

Scoring Categories and Response Frequencies

Category no.	Interaction process analysis definition	"Psychotherapeutic" definition*	Psychiatrists		Psychologists		Social workers	
			n	%	*n*	%	*n*	%
1	*Shows solidarity,* raises other's status, gives help, reward.	Gives reassurance, encouragement, shows compassion, tenderness.	41	4.0	13	5.3	35	10.1
2	*Shows tension release,* jokes, laughs, shows satisfaction.	Not applicable.†						
3	*Agrees,* shows passive acceptance, understands, concurs, complies.	Shows passive acceptance, understanding, is permissive (includes silences).	184	18.1	39	15.8	70	20.3
4	*Gives suggestion,* direction, implying autonomy for other.	Proposes course of action, defines ("structures") the therapeutic situation.	65	6.4	15	6.1	16	4.6
5	*Gives opinion,* evaluation, analysis, expresses feeling, wish.	Interprets, analyzes behavior patterns; thought in process, inferential reasoning.	164	16.1	25	10.0	23	6.7
6	*Gives orientation,* information, repeats, clarifies, confirms.	Restates, clarifies, "reflects" (Rogerian response).	108	10.6	35	14.2	59	17.1
7	*Asks for orientation,* information, repetition, confirmation.	Asks factual questions, expresses lack of knowledge, uncertainty.	31	3.0	17	6.9	18	5.2
8	*Asks for opinion,* evaluation, analysis, expression of feeling.	Explores, asks for elaboration or expression of feeling (includes open-ended questions, "nondirective leads").	367	36.2	88	35.6	118	34.3
9	*Asks for suggestion,* direction, possible ways of action.	Not applicable.†						
10	*Disagrees,* shows passive rejection, formality, withholds help.	Shows passive rejection, disbelief; ignores requests or complaints; thwarts, frustrates.	44	4.3	13	5.3	5	1.4
11	*Shows tension,* asks for help, withdraws out of field.	Not applicable.†						
12	*Shows antagonism,* deflates other's status, defends or asserts self.	Shows antagonism, aggression, sarcasm, irony, cynicism.	13	1.3	2	.8	1	.3
	Total		1,017		247		345	

* This column gives examples of the kinds of responses by therapists included in the categories, which were in no way redefined.

† The absence of responses in these categories is probably a function of the experimental conditions under which responses were elicited.

statements by a seriously disturbed near-psychotic patient, suicide threats, and other transference reactions, such as blocking, negativism, requests for direct advice, and open hostility.

Procedure and Subjects

Therapists who had agreed to participate were visited and interviewed individually. Each interview lasted about one and one-half to two hours.

The respondent was presented with the experimental series of patient statements and requested to state what response, if any, he would make to the hypothetical patient. The therapist was permitted to give as many or as few responses as he chose; "silent" responses were also considered acceptable.[4] Therapists were asked to assume that the patient statements occurred early in treatment, that the problem came up for the first time, and that none of the patients were hospitalized. All responses and comments were recorded by the investigator.

---

[4] In contrast to the related study by Phillips and Agnew (4) which restricted responses by a multiple-choice answer form.

Information concerning the therapist's training, experience, theoretical orientation, etc. was collected at the end of the interview.

The subjects were 25 psychiatrists, 7 psychologists, and 9 psychiatric social workers. All claimed allegiance to psychoanalytic or neo-Freudian principles, and most of them had received training from, or reported they had been influenced by, the Washington School of Psychiatry—a powerful agent molding the "therapeutic climate" in the Washington, D. C. area.

Fifteen psychiatrists, 3 psychologists, and 7 social workers had five or more years of experience in psychotherapy and will be treated as "experienced therapists" in subsequent analyses. While such a dichotomy implies a certain arbitrariness, the distinction is quite clear cut in the case of psychiatrists: the experienced therapists were for the most part fully qualified analysts engaged in private practice of psychoanalysis, whereas the inexperienced group ($N = 10$) was made up largely of psychiatric residents with relatively little experience in therapy. The former group thus includes the greatest concentration of professional experience.

Psychologists and social workers were typically employed by local mental hygiene clinics, and in addition to psychotherapy performed diagnostic and related functions.[5]

Results

Quantification of therapists' responses was accomplished by means of Bales' (1) system of interaction process analysis, a well-known general purpose framework for describing social interactions. The system commended itself for present purposes because of (*a*) the painstaking labor that went into its development, and (*b*) its theoretical neutrality with respect to the teachings of conflicting "schools" of psychotherapy.

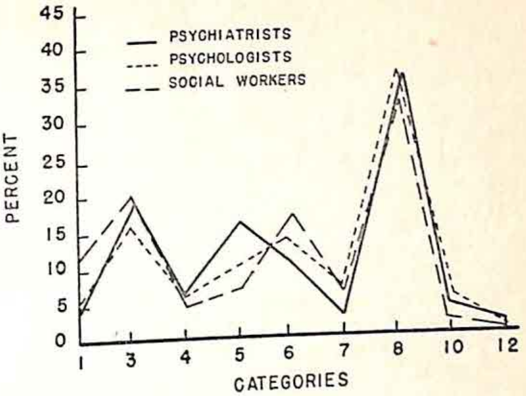

Fig. 1. Response distributions of psychotherapists by professional affiliation.

Table 1 presents Bales' 12 categories, together with the response frequencies of psychiatrists, psychologists, and social workers.

The reliability of the writer's categorizations was tested upon a stratified random sample of 10 cases comprising a total of 370 score units. Two judges who had been trained in the Bales method independently scored the ten "test" cases. Average rater agreement was 78 per cent.

Figure 1 presents the frequencies shown in Table 1 in profile form. Table 2 attempts a systematic comparison between the professional groups with respect to the percentages in any one response category. The t values (computed by the formula for uncorrelated

Table 2

Comparisons (t Values) Between Professional Groups

Response category	Psychiatrists vs. psychologists	Psychiatrists vs. social workers	Psychologists vs. social workers
1	.93	4.36**	2.09*
3	.85 P	.92	1.41
4	.18 P	1.20 P	.83 Ps
5	2.40* P	4.48** P	1.48 Ps
6	1.64	3.25**	.97
7	2.79**	1.83	.85 Ps
8	.15 P	.63 P	.35 Ps
10	.67	2.42* P	2.79** Ps
12	.63 P	1.50 P	.83 Ps

P—The response frequency of psychiatrists exceeds that of the other group.

Ps—The response frequency of psychologists exceeds that of the other group.

* Significant at the .05 level using a two-tailed test.

** Significant at the .01 level using a two-tailed test.

---

[5] A systematic comparison between the techniques of these psychoanalytically oriented psychologists and a comparable group of Rogerians has been presented previously (6).

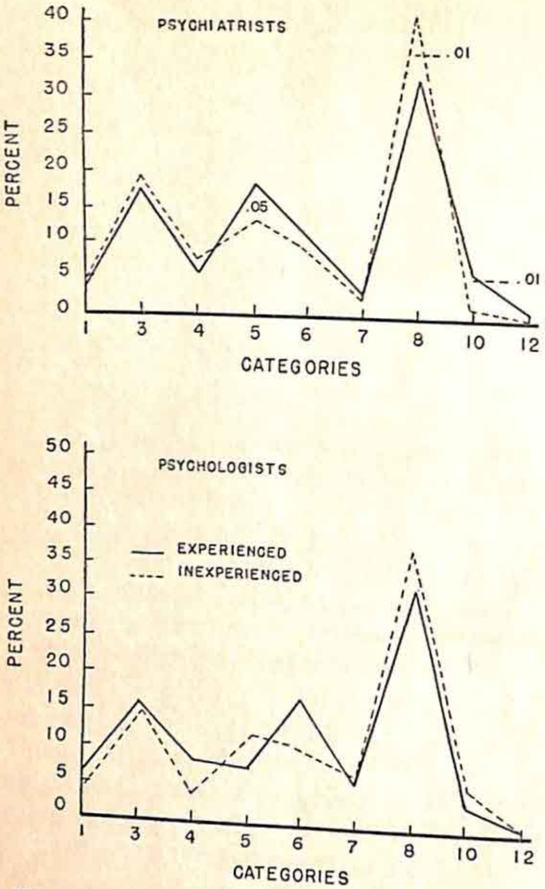

Fig. 2. Response distributions of experienced and inexperienced psychotherapists. (Experienced psychiatrists $N = 15$, $n = 595$; inexperienced psychiatrists $N = 10$, $n = 422$; experienced psychologists $N = 3$, $n = 117$; inexperienced psychologists $N = 4$, $n = 130$.)

percentages) disclose the following statistically significant differences:

1. Psychiatrists tend to give a larger number of interpretive responses (Category 5) than either psychologists or social workers.
2. Psychiatrists and psychologists exceed psychiatric social workers in passive rejections (Category 10).
3. Psychiatric social workers make greater use of reassurance (Category 1) than either psychiatrists or psychologists.

Inconsistent differences in Categories 6 and 7, while possibly quite real, have been disregarded for the present.

The next analysis addresses itself to the question: Are there significant differences between the response distributions of experienced and inexperienced therapists within a professional group? In other words: What is the effect of length of professional experience upon technique?

Figure 2 shows the results in profile form, significant differences being indicated in the graph by the appropriate significance level. Social workers, while showing a response pattern similar to psychologists, have been omitted because of the very small N for inexperienced practitioners.

The distributions are again rather similar; however:

1. Experienced psychiatrists use more interpretations and a larger number of passive rejections.
2. Inexperienced psychiatrists reveal a predilection for exploratory responses (Category 8).

With respect to psychologists, none of the differences are statistically significant, but the aforementioned shift in Category 8 seems to be in the same direction.

Discussion

The most striking finding emerging from the comparison between psychotherapists of different professional affiliations but comparable theoretical orientation is the large degree of similarity of their response profiles.

Of the three statistically significant differences between professional groups, only one (reassurance) appears to be attributable to professional affiliation. The second (interpretation) is more adequately accounted for on the basis of experience level, and the third (passive rejection) is somewhat in doubt.

1. One may regard the preponderance of reassuring responses by social workers as a move away from insight-producing "uncovering" therapy in that this technique assuages but actually changes little. This is not to say that a certain amount of reassurance may not be desirable, beneficial, or useful in the early stages of therapy, particularly with a seriously disturbed or very anxious patient.

2. The finding that psychiatrists exceed psychologists and social workers in the number of passive rejection responses is a provocative one, but because of the relatively small frequencies it is merely reported as suggestive of a possibly important difference (see below).

With regard to experience level, it is interesting to note that psychiatrists show somewhat greater divergence than psychologists. This may be a function of the greater differential in experience between experienced analysts and psychiatric residents than between psychologists of varying degrees of professional experience. While the sample of psychologists was small, the group appears to be fairly homogeneous as long as we disregard Rogerian therapists whose techniques, as we have shown elsewhere (6), are radically different.

The principal technique differences between experienced and inexperienced psychiatrists concern (*a*) interpretations, (*b*) explorations, and (*c*) passive rejections.

a. As a psychiatrist becomes more experienced, he appears to place greater emphasis upon interpretations. It was also demonstrated (5) that this characteristic not only differentiates reliably experienced psychiatrists from inexperienced ones but also from psychologists and social workers.

The significance of this result is twofold: (i) it attests to the emphasis given to interpretations by experienced psychiatrists in that this professional group follows, perhaps more closely than any other, psychoanalytic doctrine, which extols interpretation as the hallmark of intensive psychotherapy; (ii) it contravenes to some extent the recommendations of such authorities as Fromm-Reichmann (2) who caution against interpretations before sufficient information to document this therapeutic maneuver has been obtained. Thus, psychiatrists, as the most thoroughly trained group of therapists, might be expected to be more wary of interpretations, but if such cautiousness does exist, it is not revealed by the present data.

b. Concerning the statistically highly significant difference between experienced and inexperienced psychiatrists in Category 8, the following observations appear warranted. Inexperienced therapists display a tendency to ask more exploratory questions, which by and large are made at the expense of interpretations and passive rejections. It appears altogether reasonable for a therapist when confronted with a patient about whom he knows little to ask for further information; but here we have objective evidence that inexperienced psychiatrists reveal a more pronounced tendency to develop a patient's theme along exploratory lines whereas experienced psychiatrists proceed more readily to interpretations. While we cannot answer the question as to which is more effective therapeutically, there is at least the suggestion that to the extent that explorations are a sign of inexperience, less experienced therapists with medical training subscribe to the precept: If in doubt, ask for more information.

This result gains further meaning in the light of the comparison between experienced and inexperienced Rogerian therapists presented in the first study of this series (6). It was shown there that inexperienced Rogerians give a significantly larger number of reflective responses than their more experienced colleagues, the reflective response being the most favored variety. In the present study, Category 8, the preferred response of psychiatrists, reveals the same relationship. The implication is clear: Regardless of whether we deal with the Rogerian or the psychoanalytic frame of reference, the most popular (stereotyped?) technique derives a good deal of its popularity from inexperienced therapists, and undergoes a pronounced decline as the therapist becomes more experienced. It is true that this trend is more pronounced for Rogerian psychologists than for analytically oriented psychiatrists, but the same conclusion seems justified: An increase in experience leads to a diversification of therapeutic technique, and, conversely, reliance on one specific technique is a characteristic of inexperience.

c. Because of the small frequencies, it is difficult to say whether passive rejection is a technique preferred by experienced psychiatrists or by psychiatrists in general, although we are inclined to accept tentatively the former interpretation.

Summary and Conclusions

This is the second in a series of investigations to elucidate the psychotherapist's contribution to the treatment situation. It attempts a systematic comparison between the techniques used in early interviews by psychiatrists, psychologists, and psychiatric social workers subscribing to psychoanalytic principles. Therapeutic responses were secured from 25 psychiatrists, 7 psychologists, and 9 psychiatric social workers of varying degrees of professional experience by presenting a series of 27 patient statements extracted from actual therapeutic interviews. A total of 1,609 response units were categorized by Bales' system of interaction process analysis. Average rater agreement was 78 per cent. Salient results may be summarized as follows:

1. The response profiles of the three professional groups manifest a considerable degree of similarity. Of the statistically significant differences, only one appears to be a function of professional affiliation, namely the psychiatric social workers' predilection for reassurance.

2. Comparisons between experienced and inexperienced practitioners among psychiatrists and psychologists (social workers were omitted because of a small *N* for inexperi-

enced workers) disclose the following reliable differences: (*a*) experienced psychiatrists give a larger number of interpretive responses; (*b*) inexperienced psychiatrists show a preference for exploratory responses; (*c*) experienced psychiatrists use more passive rejections.

It is concluded that professional affiliation exerts a relatively minor influence upon the kinds of techniques used by psychiatrists, psychologists, and psychiatric social workers. As long as the variable of theoretical orientation is held constant, all therapists adhering to psychoanalytic principles employ very similar techniques. When experience level is taken into account, psychologists emerge as a somewhat more homogeneous group in contradistinction to psychiatrists for whom more conspicuous differences between experienced and inexperienced therapists are in evidence. Intensive training in psychotherapy seems to stress interpretations at the expense of exploratory (probing) responses and leads to a general diversification of technique. Previous evidence from Rogerian therapists provides added support for this latter conclusion.

Received July 10, 1954.

References

1. Bales, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley Press, 1950.
2. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univer. of Chicago Press, 1950.
3. Mowrer, O. H., *et al. Psychotherapy: theory and research.* New York: Ronald, 1953.
4. Phillips, E. L., & Agnew, J. W. A study of Rogers' "reflection" hypothesis. *J. clin. Psychol.*, 1953, 9, 281–284.
5. Strupp, H. H. An empirical study of certain psychotherapeutic operations: an exploration of the verbal response techniques of psychiatrists, psychologists, and psychiatric social workers. Unpublished doctor's dissertation, George Washington Univer., 1954.
6. Strupp, H. H. An objective comparison of Rogerian and psychoanalytic techniques. *J. consult. Psychol.*, 1955, **19**, 1–7.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Adjustment Changes in the Absence of Psychotherapy[1]

Rosalind F. Dymond
Counseling Center, University of Chicago

It has often been argued that there is no evidence apart from the feelings and impressions of psychotherapists themselves that psychotherapy produces changes that are any "greater" or "different" from those which would have resulted in about the same proportion of cases if they had not been seen by a psychotherapist. It is common knowledge that many people in severe psychological distress "pull out of it" or "spontaneously remit" without having been so treated.

Procedure

In the research program evaluating client-centered psychotherapy (6) some attempt was made to study this phenomenon. Nineteen of the clients who applied for psychotherapy at the Counseling Center of the University of Chicago were required to wait before beginning their interviews. These subjects were given a battery of psychological tests when they first applied and were retested with the same battery after a two-month waiting period. Admittedly two months is much too short a time to really test the possibility of spontaneous remission. It was selected because this was the average period that the non-research cases had to wait for a therapist. It was felt that it would be unfair to people in trouble to put off their therapy longer than this for research purposes when they felt the need for immediate help.

After waiting for the two-month period, 11 of these people went on to complete their psychotherapy (six or more interviews), and were tested again with the same battery for the third time. The remainder of the group, eight people, came a time or two and then failed to return. These we came to call the "attriters."

Grummon (5) has reported that there are some very interesting differences between the tests done before and after the waiting period for these two groups. He found that those who waited and then completed therapy showed no improvement over the waiting period on a variety of tests, but those who waited and then failed to get started in therapy before breaking off showed that significant improvement had taken place during their waiting period.

This finding raises a number of questions:

1. In what respects did the attrition group change and the therapy group fail to change?
2. Were the two groups drawn from the same population or were they differentiable to start with?
3. Were the changes in the attrition group of the same kind as those which are characteristic of successful psychotherapy subjects?

Results

Questions 1 and 2 will be taken up first, and the remainder of this paper will be concerned with question 3.

Table 1 shows that in terms of formal characteristics the attriters do not appear to be markedly different from the therapy subjects in terms of age, sex, and occupation (student or nonstudent).

The tests on which the attrition group showed the most marked improvement were those which involved self-description. The most dramatic of these was the *Q*-adjustment score (2). This was the scoring for adjustment of each subject's sorting of 100 self-referent items along a continuum of "this statement is very like me" to "this statement

---

[1] This investigation was supported (in part) by a research grant (PHS M903) from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

Table 1

Wait Group Subjects

	Age		Sex		Occupation*		Number of interviews	
Group	Mean	Range	M	F	S	NS	Mean	Range
Therapy ($N=11$)	27.4	22–36	7	4	8	3	36.6	6–108
Attrition ($N=8$)	24.6	20–29	3	5	3	5	2.1	1–5

* S = student; NS = nonstudent.

is very unlike me." The basic instrument was developed by Butler and Haigh (1). Table 2 gives the prewait and pretherapy scores of the two groups.

The difference between the mean scores of the therapy group at the two testing points was not significantly different using t for correlated means, but for those who did not go into therapy the difference between the prewait and pretherapy means was significant at the 2% level, using the same statistic. The difference between the means of the two groups at the first testing was not significant. These data seem to support the conclusion that, although the groups were about equally maladjusted initially as self-described, eight of them managed to improve while waiting for therapy and then chose not to continue their interviews. The 11 who did not improve while waiting for therapy to begin, as measured from their self-description, did go on to complete therapy. However, all these conclusions are based on what the individual is able or willing to say about himself.

Another instrument which formed part of the battery administered before and after the two-month wait was a standard 20-card Thematic Apperception Test (TAT). The prewait and pretherapy records of the 19 subjects were mixed in with the records of all the other subjects in this experimental program (6). They were coded and all identifying material was removed. A mental health rating was then made for each record as a whole (4). Each record was rated from 1 to 7 with a rating of 1 being equivalent to "bordering on psychotic or psychotic," and 7 being a "well-integrated, happy person, socially effective." Table 3 compares the two wait groups in terms of their TAT ratings.

Table 2

Q-Adjustment Score of Therapy and Attrition Groups

	Prewait		Pretherapy	
Group	Mean	Range	Mean	Range
Therapy	28.6	16–38	30.2	15–41
Attrition	32.0	22–46	42.9	25–55

Here, none of the t's reached significance, showing that the groups were equally maladjusted at the initial testing point and neither improved significantly over the waiting period. This result is sharply at variance with the results of the *Q*-adjustment measure. What is the relation of these two adjustment scores? Rho correlations were run between the TAT ratings and the *Q*-adjustment scores for both groups at the two testing times. The rho's for the therapy group were + .47 at prewait and + .30 at pretherapy. For the attrition group, the rho at prewait was + .39 but at pretherapy it was − .20! These results and those of a previous study (3) would seem to show that although there is ordinarily a fair relationship between these two measures of adjustment, this relationship does not hold for the attrition group at pretherapy time. This seems to cast some doubt on the apparent improvement in the waiting period of these people.

Table 3

TAT Mental Health Ratings of Therapy and Attrition Groups

	Prewait		Pretherapy	
Group	Mean	Range	Mean	Range
Therapy	3.0	1–4	3.4	1–5
Attrition	3.6	3–5	3.7	3–5

The lack of support from the TAT analysis for the positive adjustment change in self-description raises the question: Are the self-descriptive changes defensive in nature, perhaps in response to the threat of therapy? The question brings us to question 3: Of what does the improvement in the absence of

Table 4

Comparison of Attrition and Success Groups on *Q* Adjustment and TAT Ratings

Group	*Q* adjustment			TAT rating		
	First score	Second score	Change	First score	Second score	Change
Attrition ($N = 6$)			+14.37**			0
Mean	33.0	47.3		3.5	3.5	
Range	22–46	40–55		3–4	3–5	
Success ($N = 6$)			+13.50**			+1.2*
Mean	34.83	48.33		3.8	5.0	
Range	23–42	35–54		3–5	4–6	

Note.—For the attrition group the first score was their score at prewait and improvement was change from prewait to pretherapy, while the success cases' first score was established at pretherapy and improvement was change from pretherapy to posttherapy.

* Significant at 5% level.
** Significant at better than 1% level.

therapy consist? Is the improvement in the attrition group during the no-therapy period of the same kind as that which results from a period of successful psychotherapy? The pattern of large *Q*-adjustment–score increase unsupported by commensurate change in the TAT rating did not hold for all eight of the attriters to an equal extent. However, for six of them the pattern was clear. These six were matched with a group of successful therapy cases in terms of their initial status and degree of improvement on the *Q*-adjustment score. "Success" was established by a threefold criterion: (*a*) *Q*-adjustment–score increase from pretherapy to posttherapy, (*b*) TAT-rating increase from pretherapy to posttherapy, and (*c*) the therapist's rating of the case on a nine-point scale of improvement at six or above (7). The matching data are shown in Table 4.

To make a start on the problem as stated in question 3 above, the original self-descriptive *Q* sorts were examined. Each of the subjects had sorted the 100 self-referent items at the different testing points into nine piles under the instructions, "Sort the cards to describe yourself as you see yourself today, putting the card that is least like you at the 'zero' end and the one that is most like you at the 'eight' end, then the four cards that are next most and least like you at 'one' and 'seven' and so on." The required distribution is shown in Table 5.

Table 5

Q-Sort Distribution

	Least like me								Most like me
Pile number	0	1	2	3	4	5	6	7	8
No. of cards	1	4	11	21	26	21	11	4	1

Each item was given the weight of the pile number into which it fell in the particular sorting. The average sorting that the six attrition subjects made at the pretherapy point was then computed and compared to the average sorting made by the six matched success cases at their posttherapy point. The differences between the two groups in the mean weighting of each item were tested for significance using *t*. Table 6 lists the items which were significantly different with respect to their average weighting.

Although the smallness of the *N*'s on which these results were based puts some stricture on the degree of interpretation they warrant, the picture of the attrition group that emerges from a reading of the items is so clear that it seems probable that something like it would be found in a study with more adequate samples. The attriters appear to be a group of "rugged individualists" who are hanging on tight to the idea, "I *am* someone, I *am* superior, I *do* want to cope and will do it *myself*." They seem to have difficulty accepting the idea of help from others (I am an aloof, reserved person; I make strong demands on myself; I am self-reliant, competitive; I have

Table 6

Q Items Which Differentiate Attrition and Success Groups

Item	Mean weighting		p
	Success	Attrition	
Weighted *more* characteristic by attrition group			
I am an aloof, reserved person.	2.3	4.3	.01
I am ambitious.	4.3	6.0	.02
I feel superior.	3.5	4.6	.02
I am self-reliant.	4.3	5.5	.05
I am afraid of sex.	2.0	3.5	.05
I just don't respect myself.	2.1	3.1	.05
I am a competitive person.	3.6	4.8	.10
I make strong demands on myself.	3.6	5.0	.10
I don't trust my emotions.	3.0	4.5	.10
I have to protect myself with excuses, with rationalizing.	3.1	4.3	.10
Weighted *less* characteristic by attrition group			
I want to give up trying to cope with the world.	3.5	1.1	.01
I am no one. Nothing really seems to be me.	3.8	1.6	.02
I am liked by most people who know me.	6.8	5.6	.05
I understand myself.	6.0	4.3	.05
I express my emotions freely.	5.6	4.0	.10

to protect myself with excuses, with rationalizing). It does not seem surprising in the light of this pattern that they have made a strong effort during the waiting period to pull themselves up by their own bootstraps; they have decided that therapy is not for them and do not follow it through; and their TAT's do not show any deep personality change.

What are the characteristics of the successful psychotherapy subjects in contradistinction to the attriters who "go it alone"? Some answer is given by the items: I express my emotions freely, I am liked by most people who know me, I trust my emotions, I respect myself, I understand myself. All of these were more characteristic of the self-description of the successes than of the attriters.

Summary

This study has been a bare beginning on a question which has been an open one far too long. Although there are many shortcomings in the study presented here, some conclusions can be drawn. A summary may help to clarify the points to be made.

1. There was a small group of six subjects who appeared to have "spontaneously recovered," that is, decided after waiting for therapy that they did not want therapy after all, and appeared to have improved in adjustment, as measured from their self-description, about as much as those who went through therapy successfully. This group was not initially differentiable in terms of adjustment status from the subjects who did *not* improve in adjustment while waiting and who *did* go into therapy.

2. The attriters' increase in adjustment as measured from the self-descriptive *Q* sorts was not confirmed by a commensurate increase in mental health as rated blindly from their TAT's. On the basis of the TAT ratings, "the spontaneous remission" group appeared to be as poorly adjusted at the end of the waiting period as they were when they first applied for therapy, not significantly different in adjustment from those who then continued with therapy, and significantly poorer in adjustment than those who completed therapy. In other words, in terms of the adjustment scoring of the subject's own self-description the attriters had improved, whereas in terms of the scoring of their projective tests by a diagnostician they had not.

3. To discover whether the "improved" self-description of the attriters was differentiable from the improved self-description of a group which had completed therapy successfully, the attriters were matched with a successful therapy group in terms of initial status and degree of improvement on the *Q*-adjustment measure. From a comparison of the average "improved" self-description of these two groups, 15 items were found which were weighted in terms of "like me" or "unlike me" significantly differently. A study of these crucial items appeared to show that the "improved" self-description of those who did not go into therapy had strong defensive and compulsive elements which were absent or significantly less characteristic of the self in successful cases as described by them.

This study, limited as it is by the small number of cases and short period of time covered, has shown that although positive adjustment changes appear to take place in maladjusted persons in the absence of psychotherapy, these are not identical to the changes which occur in equally maladjusted persons who complete therapy. No "deep" reorganization appears to take place. The "improvement" appears to be characterized by a strengthening of neurotic defenses and a denial of the need for help.

In future studies on this subject, not only should the sample be enlarged and the period of no therapy be extended, but more data should be gathered: additional diagnostic data to discover the nature of the defense structure of those who do and do not continue with therapy, and follow-up data so that the effects of self-improvement of this kind may be compared to the effects of successful psychotherapy.

Received July 16, 1954.

References

1. Butler, J. M., & Haigh, G. V. Changes in the relation between self-concepts and ideal concepts. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 55–75.
2. Dymond, Rosalind F. An adjustment score for *Q* sorts. *J. consult. Psychol.,* 1953, **17**, 339–342.
3. Dymond, Rosalind F. Adjustment changes over therapy from self-sorts. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 76–84.
4. Dymond, Rosalind F. Adjustment changes over therapy from Thematic Apperception Test ratings. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 109–120.
5. Grummon, D. L. Personality changes as a function of time in persons motivated for therapy. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 238–255.
6. Rogers, C. R., & Dymond, Rosalind F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.
7. Seeman, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 99–108.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Language and the Interpretation of "Sexual" Symbolism[1]

Harold H. Mosak
Roosevelt University

In current psychoanalytic practice the interpretation of symbols is accorded two types of treatment. Some psychoanalysts believe that certain symbols are fixed symbols while others regard these symbols as meaningful only when interpreted in the context of the individual or group using these symbols. Freud advocated the first position in his statement that "there are symbols which preponderantly, or almost exclusively, designate one of the sexes, and there are yet others which . . . have only the male or only the female signification. To use long, stiff objects and weapons as symbols of the female genitals, or hollow objects . . . as symbols of the male genitals is not permitted by the imagination."

Freud's statement would lead us to expect that language, also a symbolic form, should bear the stamp of these unconscious processes. Consequently, a list of "sexual" symbols was translated into three languages—French, Spanish, and German—which designate gender for nouns. If Freud's statement is valid, we might assume that symbol and gender designations would coincide. Our results indicate that Freud's contention is not borne out in language. In French 15 of the 33 "male" symbols possess feminine gender while 12 of the 26 "female" symbols are designated by masculine articles. Spanish assigns feminine articles to 19 male symbols and masculine articles to 12 female symbols. In German symbol-gender reversals occur for 11 of the male symbols and 9 of the female symbols. We also note that the designation of "long, stiff objects and weapons" as feminine and "hollow objects" as masculine is indeed "permitted by the imagination." No consistency in gender appears for the three languages. Only 21 of the 59 symbols are consistently used with the same gender, and of these 21, 8 demonstrate consistent symbol-gender reversals. These results warrant seriously questioning whether these symbols are either fixed or sexual. It might be argued that Freud's theory of symbols was based on his expressed feeling that dreams and dream elements are preponderantly visual images and therefore would not possess parallelism with language symbolism. Freud himself confutes such an argument when he uses etymological evidence to corroborate his theory of dream symbolism.

In the area of psychotherapeutic investigation such results as ours confront us with an anomalous situation. For despite lack of theoretical foundation or experimental verification, therapeutic techniques may nevertheless prove effective. This effectiveness might even be utilized to demonstrate the *validity* of the assumptions underlying the use of the technique. Actually, this effectiveness merely demonstrates the *utility* of the technique. In the framework of psychoanalytic theory, Freud's hypothesis might possess such utility. For other schools of psychotherapy the theory of fixed sexual symbols would presumably possess little utility or validity.

Brief Report
Received November 22, 1954.

---

[1] An extended report of this study may be obtained without charge from Harold Mosak, 4305 W. Flournoy St., Chicago, Ill., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4466 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Conformance in Role Perception as a Personality Variable[1]

Theodore R. Sarbin and Curtis D. Hardyck
University of California, Berkeley

This study, one of a series growing out of a social psychological theory of behavior (5), attempts to clarify "conformance," a perceptual variable, and throw some light on its personological correlates.

Conformance is defined as the modal perceptual response to a social stimulus object. It is parallel to the notion of conformity as applied to overt conduct. When a social object elicits qualitative perceptual responses from members of a specified subculture, a relatively large proportion of the reference population will agree on the quality represented by the object; a small number will differ from the majority but will agree among themselves; and a few will agree with virtually no one. Figure 1 illustrates the notion graphically.

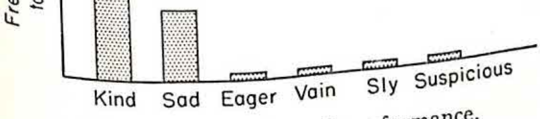

Fig. 1. An illustration of conformance.

The stimulus material employed in this study was the same as that developed for a taxonomy of postural cues used in role perception and inference (6). We prepared a set of 44 schematized "stick figures" of humans in various postures. The figures were drawn by an interpretive dancer and represent a wide range of expressive and attitudinal states. The figures were so constructed as to offer no cues for interpretation other than posture. Figure 2 is one of the items in the Stick Figures Test.

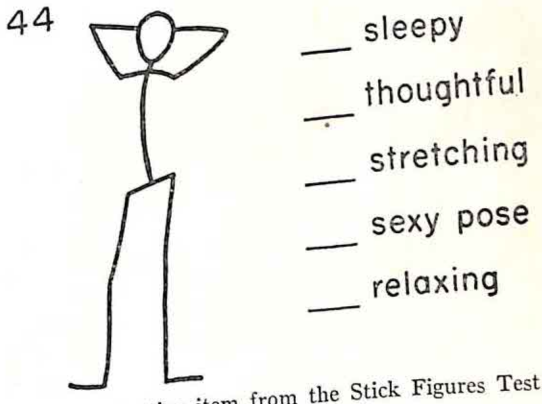

Fig. 2. Illustrative item from the Stick Figures Test.

As a preliminary step the figures were presented on 3 × 5 lantern slides to a group of 150 upper-division psychology students. The students were given the following instructions:

> A number of stick figures, representing people in various postures, will be shown to you. As the figures are shown, you are to write 1 or 2 adjectives which best describe your impression of that figure.

From the data obtained in this sample, a check list of five adjectives, ranging from most frequent to least frequent, was constructed for each figure. As no significant sex differences were found, men and women were combined in preparing the check lists. The reliability of the adjective lists was determined by taking a second sample of 70 upper-division psychology students and repeating the procedure given above. No significant differences in proportion or kind of adjectives were found.

---

[1] No. 7 in a series entitled "Contributions to Role-taking Theory." The authors wish to express their gratitude to Professor H. G. Gough, Drs. Frank Barron, Albert Kostlan for serving as judges.

In developing norms for the use of the stick figures, we noted impressionistically that certain persons characteristically selected nonmodal, i.e., nonconformant responses. From these impressions and from certain implied postulates of role theory, we were led to the following hypothesis: persons who frequently make nonconformant perceptual responses are characterized by social behavior which is deviant or socially invalid.

The theoretical basis for this hypothesis follows from notions originally advanced by G. H. Mead (4) and more recently developed by the senior author (5). Role perception is the first part of a social act; it is the (usually) silent naming or locating of the position of the other on the basis of available cues. This serves as an aid in locating the position of the self. The motoric response, the role enactment, follows from the perceptual response. Since the perceptual and motoric aspects of the social act are acquired holistically, if the perceptual response is conformant, then there is a high probability of socially valid (though not necessarily conforming) role enactment. Thus, the veridicality of one's role perceptions bears a positive relationship to functional adaptation in the social world.

Following are reports of three experiments which were carried out to test the hypothesis.

Schizophrenic vs. Normal Responses

As an initial check on the validity of our procedure, the stick figures were administered individually to 30 cooperative, undeteriorated, hospitalized schizophrenics. They were given the following instructions:

> A number of stick figures, representing people in various postures, will be shown to you. Each figure will be numbered, corresponding to a number on this paper. For the corresponding number on this paper, there is a list of five words, plus a sixth space which is left blank. As the figures are shown to you, you are to check the word which you think best describes your impression of that figure.
>
> If the feel that none of the words on the list adequately describes your impression of the figure, write what you consider to be the best description in the space marked "other."

No time limit was set, and the experimenter was allowed to rephrase the instructions if a subject failed to comprehend the initial instructions.

The distributions of responses for the schizophrenic sample were compared with the norms obtained on our previous samples. For the schizophrenic sample the distributions were all uniformly rectangular, in comparison to the **J** curves obtained from college normals, thus offering substantial evidence for the usefulness of the technique and for the initial confirmation of the hypothesis.

Conformance Scores and Social Adjustment

In the second exploratory study, because of practical limitations, we were unable to obtain direct behavioral rating such as would be obtained in an assessment program. Instead we made use of indirect sources of information on role enactment. The California Psychological Inventory (Gough, 2) has a number of scales, all of which are empirically rooted in social behavior. This instrument was administered to 21 undergraduates who had been subjects for the Stick Figures Test. The Inventory profiles were then ranked by Gough according to a "social adjustment" criterion.

Responses to the stick figures were scored as follows: The modal response to each stick figure for the previous samples was assigned a weight of 1. The subjects were then given a score in terms of the number of modal response choices made. These scores were ranked and a Spearman rho was computed between the stick figure score rankings and the ranking on social adjustment. The correlation was .81 ($p = .01$), a highly significant relationship considering the small number of subjects. On the basis of the second preliminary study, our hypothesis would seem to be further supported. However, a cross validation was indicated owing to the small N and the use of a single judgment of social behavior.

Cross Validation of Conformance and Social Adjustment

The next step involved the presentation of the figures to a sample of 91 upper-division psychology students. The instructions and scoring method were the same as in the two previous studies. For this group the Minnesota Multiphasic Personality Inventory (MMPI) was used as a measure of personality functioning. The MMPI profiles were sorted by each of three judges into a normal distribution on the following criterion of *Adjustment:*

> Getting along well in the world as it is. Adequate degree of social conformity; capacity to adapt to a wide range of conditions. Ability to fit in (3).

The results were as follows: The range of scores on the Stick Figures Test was from 11 to 26 (44 stick figures). The correlations of the conformance scores on the Stick Figures Test with the sorts of the MMPI profiles were .32 for Judge A, .68 for Judge B, and .52 for Judge C.

Combining of the three sorts to obtain a more stable criterion of social adjustment resulted in a Pearsonian product-moment correlation of .51 with the conformance score. The correlation is unusually high, especially in the absence of a refined method for obtaining the conformance score.[2] The above sorts are done with sex groups combined. Separate sorts for males and females revealed no significant differences among the correlation coefficients.

Interjudge reliabilities on social adjustment were .66 between Judges A and B, .73 between A and C, and .83 between B and C. They are considered adequate for our experimental purposes. The split-half reliability for the Stick Figures Test is .50. Refinement of the test by item analysis will doubtless increase the reliability.

Correlations between conformance score and certain clinical scales provide some additional confirmatory evidence. The F scale on the MMPI indicates the number of "infrequently answered" items in the test. In a sense it is a subtle measure of conformance. Pearsonian r is $-.32$. Subjects who score high on the F scale tend to preceive the postures in a nonconformant way. Another index of social maladjustment is the number of clinical scales on the MMPI on which the subject's score is two sigmas or more above the mean. This index correlates with conformance $-.33$ (and was probably used implicitly by all the judges in making their judgments of social adjustment). An empirically devised scale of ego strength (1) correlates .50 with conformance.[3]

Summary

On the basis of certain postulates derived from role theory, a significant relationship between nonconformant perceptual responses and deviant social behavior is hypothesized. Three experiments were performed to test this hypothesis. The first experiment compared the distribution of responses of normals and schizophrenics to a set of schematized stick figures of humans in various postures. The normal subjects were found to have J-curve distributions as contrasted with the rectangular distributions of the schizophrenic sample.

In a second experiment, the number of conformant responses of a group of normal subjects was found to bear a significant relationship to their social adjustment as judged by test profiles. A third, cross-validational, experiment indicated the stability of the relationship. Correlations with certain empirically derived personality scales offer further confirmation of the hypothesis.

Received July 12, 1954.

References

1. Barron, F. An ego-strength scale which predicts response to therapy. *J. consult. Psychol.*, 1953, 17, 327–333.
2. Gough, H. G. *The California Psychological Inventory*. Berkeley: Univer. of California Press, 1951.
3. Gough, H. G. Predicting success in graduate training: a progress report. Unpublished manuscript, Univer. of California, Institute of Personality Assessment and Research, 1950.
4. Mead, G. H. *Mind, self and society*. Chicago: Univer. of Chicago Press, 1934.
5. Sarbin, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 223–258.
6. Sarbin, T. R., & Taft, R. *An essay on inference in the psychological sciences*. Berkeley: Garden Library Press, 1952.

---

[2] In view of the method used to obtain the conformance score, the possibility of curvilinearity should be taken into account. Scatter plots of the correlations revealed no significant tendency toward curvilinearity. It would seem safe to conclude, therefore, that judgments in terms of conformance do bear a significant relationship to level of adjustment as measured by the MMPI profile sorts.

[3] Conformance scores and an indirect measure of kinesthetic inference (empathy) are correlated to the extent of .48. This finding is only peripherally related to the hypothesis and is not discussed in the present paper.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Rorschach Experience Balance in Epileptics

Paul W. Pruyser[1] and Angela T. Folsom
The National Veterans Epilepsy Center, Boston

The Rorschach literature contains several observations of abnormalities in the use of the human movement and color determinants by epileptics. Rorschach (9), using mostly deteriorated epileptics, reported an underproduction of *M* in slowly dementing patients and a marked overproduction of *M* in the most demented ones and concluded that the quantity of *M* in an epileptic record is proportional to the degree of dementia. Rorschach described an overemphasis on the color determinants, also increasing with the degree of dementia, and often formed by the use of many pure *C* responses, including the so-called color-naming (*Cn*) response. Since the publication of Rorschach's manual several workers have partially or wholly corroborated his findings, except for his observation that the degree of *M* increases with more pronounced dementia. In intramural as well as extramural groups of epileptics the finding of low *M* is general (1, 4, 6, 7, 10, 12, 13) and several workers have confirmed the notion of a typically epileptic, extratensive experience balance (EB). A disproportionate emphasis on color, usually with predominance of pure *C* and *CF* over *FC* has been found to contribute to the extratensive tendency (1, 4, 10, 12). These indices have also been used to validate the old, widespread notion of the so-called "emotional explosiveness" of epileptics and have even served as an aid in diagnosing epileptics from Rorschach records (5). As a matter of fact, it appears that Rorschach's own use of the findings of the predominance of unformed color in his group of epileptics played a role in defining the diagnostic meaning of the *CF* and *C* responses as indicating emotional explosiveness (9, Tables I–VII and XII). Obviously, this part of the validation of color responses is based upon a circular argument which starts with a clinical notion and interprets the meaning of an abnormal test response directly in terms of that clinical conviction even before it is systematically verified.

Studies of the intellectual functioning of epileptics have shown that samples from different strata of the epileptic population give very different results. For instance, whereas testing of institutionalized patients shows on the whole a subnormal range of IQ's, samples of extramural patients have yielded a mean IQ and a range comparable to those of the normal, nonepileptic population (3, 13). It is reasonable to assume, therefore, that the selection of samples may also affect the Rorschach findings, the more so since it seems now well established that the frequency of *M* responses is positively correlated with IQ level. To what extent the normal correlation between IQ and amount of *M* holds for epileptics is as yet unclear, since accurate data on IQ's are lacking in several of the earlier Rorschach publications. With few exceptions (6, 7, 13), most authors have published on deteriorated and/or institutionalized groups, or on samples with an abnormal proportion of below average IQ's.

Present Study

The present study offers EB data on a group of 136 male epileptic veterans, all extramural patients, who were hospitalized only for diagnostic study and medical treatment of their symptoms. They are probably a representative group of intellectually grossly intact epileptics. They have a mean Wechsler-Bellevue Full Scale IQ of 102.6 with a range from 78 to 129; their mean age is 31 (range 21–52). In the vast majority the epileptic symptoms made their first appearance after adolescence. The group permits the following subdivisions:

---

[1] Now at Topeka State Hospital, Topeka, Kansas.

22	with idiopathic grand mal epilepsy
15	with idiopathic psychomotor epilepsy
23	with idiopathic grand mal and psychomotor epilepsy
39	with symptomatic grand mal epilepsy
9	with symptomatic psychomotor epilepsy
28	with symptomatic grand mal and psychomotor epilepsy

In order to assess the generality of the previously reported abnormalities in the EB of epileptics, which is the purpose of this study, we propose:

a. a statistical comparison of our epileptic group with a nonepileptic group of normal intellectual standing;

b. a more detailed analysis of the components of the EB.

Results

The publication of Rorschach norms by Beck *et al.* (2), based upon the Spiegel sample, provides a useful frame of reference for statistical comparison. Table 1 shows the results of a comparison between Beck's group of normals and our group of epileptics for the determinants which constitute the experience balance. In order to cope with the gross deviations from the normal distribution curve which characterize most Rorschach scores the following procedure was adopted: a median score was computed for the two groups combined[2] and this score was taken as the cutting point for the development of a 2 × 2 table containing for each group the number of cases falling below and above that point. The p values derived from a chi-square test would then indicate to what extent the null hypothesis, in this case whether the two groups could be seen as samples from the same population, was tenable. The table gives the relevant figures and indicates that significant differences exist between the two groups. For both movement and color responses the epileptic group shows a striking underproductivity, which contrasts with previously reported findings of an overuse of color by epileptics as a group.

Table 1

Comparison of EB Determinants in 136 Epileptics and Beck's Normals

Measure		N below combined group median	N above combined group median	p
M				
	Normals	46	111	.001
	Epileptics	93	43	
FC				
	Normals	44	113	.01
	Epileptics	62	74	
CF				
	Normals	55	102	.02–.01
	Epileptics	66	70	
C				
	Normals	100	57	.01
	Epileptics	109	27	
Sum C				
	Normals	47	110	.02–.01
	Epileptics	63	73	

Table 2

Frequency of Each Number of M, FC, CF, and C Responses in 136 Epileptics

Frequency	Response M	FC	CF	C
0	37	62	66	109
1	56	42	43	18
2	15	13	11	2
3	18	10	8	6
4	5	6	4	1
5	3	3	3	0
6	1	0	0	0
7	1	0	0	0
8	0	0	1	0

As an introduction to a further analysis of these data we specify in Table 2 the frequencies for each number of M, FC, CF, and C responses for the total epileptic group. Inspection of the figures shows that there is a sharp increase in the zero category if one moves from M, through FC and CF, to C. To assess these data properly and to deal with the possible influence of R, we employed the measure $M\%$ as introduced by Rapaport

---

[2] We express our indebtedness to Drs. Donald W. Fiske and Samuel J. Beck of the University of Chicago. The former was kind enough to furnish the raw scores of the Spiegel sample which enabled the use of distribution-free statistics.

Table 3

$M\%$ in 136 Epileptics and Beck's Normals

Frequency (in per cent) in each class	Group			
	Idiopathic ($N=60$)	Symptomatic ($N=76$)	All epileptics ($N=136$)	Normals ($N=157$)
0	32	24	27	15
1–20	5	5	5	2
21–40	20	18	19	18
41–60	12	22	18	23
61–80	15	17	16	24
81–100	17	13	15	18

et al. (8). It is based on the formula

$$\frac{M}{M + Sum\ C} \times 100$$

computed for each record separately, expressing in the range 0–50 a color prevalence, and in the 51–100 range a movement prevalence. Values between 40% and 60% can be taken as without clear-cut prevalence in either direction. Table 3 shows the percentage of cases falling in each of five intervals of an $M\%$ scale, with the corresponding figures of Beck's normal group for comparison. For the epileptic group as a whole, 51% of all cases seem to show color predominance, against 31% with a movement predominance. The first figure, however, includes cases in which the $M\%$ was zero as a result of total absence of M. In such cases the $M\%$ is not a fair index and should not be used; when we exclude these cases from the color-predominant group, only 24% of all cases have a clear-cut color prevalence in the range from 1 to 40. Such a low incidence of color prevalence in our group is in sharp contrast to earlier reports in the literature which show a general predominance of color responses for epileptics. Moreover, a further analysis of those cases which have an $M\%$ from 0 to 40 shows that 53% of them had no M, and 40% had only one M, whereas only 7% had two or more M. In other words, in the majority of cases with a seeming predominance of color the real abnormality lies in a marked underproduction of M rather than in any overproduction of C, CF, or FC. A comparison of the distributions of the total epileptic with the normal group by means of chi square (based on the raw data instead of the percentages shown in the table) yields a p between .02 and .05.

A more familiar approach to EB lies in assigning each case to one of the following four categories, as recently more clearly defined by Stotsky (11):

1. Coarcted: if both M and Sum C are equal to one or less
2. Ambiequal: if either M or Sum C are greater than one and the difference between them is one or less
3. Introversive: if M is greater than Sum C plus one
4. Extratensive: if Sum C is greater than M plus one

Table 4 gives the percentages of cases falling in each of these categories, as compared to the normal group. Extratensive cases form only 28% of the total group (comparable to the 26% of the normal group), whereas the coarcted and ambiequal cases, which show no prevalence in either direction, constitute together more than half of the total sample (as against only 35% of the normals). If one further takes notice of the fact that in the EB classification the extratensive group also contains a number of cases with only one M, it appears that the notion of color predominance in the Rorschach records of epileptics as a group finds no support in these data. A statistical check with the normal group by means of chi square (based on the raw data instead of the percentages shown in the table) indicates a difference significant at the 1% level, mostly due to the predominance of coarcted and the relative absence of introversive cases in our sample.

It has been said (1, 4, 9) that color predominance in epileptics takes the form of a

Table 4

EB Classification of 136 Epileptics and Beck's Normals

EB	Idiopathic %	Symptomatic %	All epileptics %	Normals %
Coarcted	35	33	34	12
Ambiequal	20	21	21	23
Introversive	18	17	17	39
Extratensive	27	29	28	26

Table 5

Comparison of Idiopathic and Symptomatic Epileptics

Rorschach factor	Idiopathic ($N=60$) Median	Symptomatic ($N=76$) Median	p
Determinant:			
M	.71	.63	.50
FC	.48	.68	.01–.02
CF	.44	.61	.10–.20
C	.33	.30	.50–.70
Sum C	.50	.64	.20
$M\%$	33.32	42.26	.50–.70
	Number	Number	
EB class:			
Coarcted	21	25	.20–.30
Ambiequal	12	16	
Introversive	11	13	
Extratensive	16	22	

disproportion within Sum *C*, consisting of an excess of pure *C* and *CF* responses over *FC*. One glance at Table 2 shows that this observation does not hold for our group of epileptics.

Others (7, 10, 12) have pointed out that many of the pure color responses may consist in simple color naming, the so-called *Cn* responses. They occurred in 10 cases or 7% of the whole group.

In view of the fact that others (10) have reported differences in various Rorschach scores for idiopathic and symptomatic epileptics, we also compared these two groups on the main measures used in this study. The results are brought together in Table 5. The frequency distributions of the two groups do not differ appreciably, except for *FC* on which the symptomatic group scores higher. In order to represent the more typical case, and to show once more the trend toward underproductivity on either side of the EB, we chose to publish the median values in Table 5; the preceding tables contain indications about the spread of the scores.

Discussion

The findings here reported are felt to throw considerable doubt upon the generality of inferences made from earlier studies which suggested an overemphasis on color, especially on pure *C* and *CF*, in epileptics. On the other hand they confirm once more the earlier findings of low *M* in epileptic records. Our analysis has shown, moreover, that in several cases the impression of an apparent emphasis on color is based on the artifact produced by a scarcity or even total absence of *M*. For the sample of epileptics we used, then, the outstanding feature of the EB is one of underproductivity on both sides.

Since our sample differs from those of many previous workers in intellectual level and range among other things, the conclusion suggests itself that the general mental status of epileptics (which largely determines institutionalization or extramural care) is an important factor in regulating the use of color in the Rorschach test. Guirdham (4) also called attention to this factor, but he still held to the finding of an imbalance within Sum *C* consisting in a relative neglect of *FC* in favor of *CF* and *C*. This phenomenon is not typical of our group.

The idea of a typical emotional explosiveness of epileptics, which seemed to have received some corroboration from earlier studies (mainly with institutionalized patients), would thus appear in need of experimental support. And such support should preferably not come from test scores with an insufficiently defined meaning in which the clinical impression itself functions as an interpretative factor.

An intellectually average group of epileptics, such as our sample, seems to yield a variety of EB types in which constrictive tendencies predominate. The observed variability ties in well with the growing conviction among epileptologists that personality disturbances in epileptics do not follow one typical pattern, but are multifactorial with a considerable tendency toward constriction. A narrowing of the scope of functioning is found in many groups of afflicted persons and is an earmark of general neuroticism.

Since coarctation on the Rorschach test may be related to *R*, the data were also analyzed in terms of a split between a low (*R*, 0–24) and a high (*R*, 25 or more) productivity group, each compared with the analogous subdivision of the normal group. The trend of the differences was found to lie in the same direction. For *M*, the difference between epileptics and normals remained significant in the high as well as in the low productivity group; for *C*, a significant difference was found in the high productivity group.

Conclusions

1. There is a significant difference in the EB of normal subjects and of an extramural, nondeteriorated group of epileptic veterans who show underproductivity in all determinants of the EB.

2. In contrast to the bulk of previous studies on epileptics, which point to a predominance of extratensive EB's, this extramural, nondeteriorated group of epileptics yields a variety of EB types in which constrictive tendencies predominate. Generalizations about "epileptics" from the Rorschach EB become thus a hazardous undertaking.

3. In many cases the impression of an apparent emphasis on color is based upon an artifact, produced by a scarcity or even total absence of movement responses in epileptics.

4. No significant differences in the crucial determinants of the EB (M, pure C, Sum C) are found between idiopathic and symptomatic epileptics (both nondeteriorated, extramural, male veterans).

5. In view of earlier publications relating aberrations in the use of the Rorschach color determinants to emotional explosiveness of epileptics, one is thus forced to select one of the following alternatives:

a. One retains the clinical notion of the emotional explosiveness of epileptics, but then the unformed color determinant of the Rorschach test does not appear to be a valid measure of this trait.

b. One accepts the diagnostic meaning of unformed color in the Rorschach test as emotional explosiveness, but then the corresponding personality trait would not appear typical of epileptics.

Summary

The Rorschach EB data of 136 extramural, intellectually grossly intact epileptic veterans were compared with those of Beck's normal controls. In all determinants of the EB the epileptics showed a significant underproductivity. The majority had a coarcted or ambiequal EB, whereas in many cases the impression of an apparent emphasis on color was based on an artifact produced by scarcity or absence of movement responses. These facts contrast with repeated statements in the literature that emotional explosiveness (as measured by an extratensive EB) is a typically epileptic character trait. It is concluded that either: (*a*) all epileptics are emotionally explosive, but Rorschach's CF and C do not measure it; or (*b*) CF and C are valid indicators of emotional explosiveness, but epileptics as a group are not characterized by this trait.

Received July 20, 1954.

References

1. Altable, J. P. Rorschach psychodiagnosis in a group of epileptic children. *Nerv. Child,* 1947, **16**, 22–23.
2. Beck, S. J., Rabin, A. I., Thiessen, W. G., Molish, H., & Thetford, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.,* 1950, **30**, 241–298.
3. Collins, A. L., & Lennox, W. G. The intelligence of 300 private epileptic patients. *Proc. Ass. Res. nerv. ment. Dis.,* 1947, **26**, 586–603.
4. Guirdham, A. The Rorschach test in epileptics. *J. ment. Sci.,* 1935, **81**, 870–893.
5. Kelley, D. M., & Margulies, H. Rorschach case studies in the convulsive states. *Rorschach Res. Exch.,* 1940, **4**, 157–190.
6. Lisanski, E. S. Convulsive disorder and personality. *J. abnorm. soc. Psychol.,* 1948, **43**, 29–37.
7. Piotrowski, Z. The personality of the epileptic: the concept of the epileptic personality. In P. H. Hoch & R. P. Knight (Eds.), *Epilepsy, psychiatric aspects of convulsive disorders.* New York: Grune & Stratton, 1947. Pp. 89–108.
8. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Vol. 2. Chicago: Year Book Publishers, 1946.
9. Rorschach, H. *Psychodiagnostics.* Berne: Huber, 1942.
10. Stauder, K. H. *Konstitution und Wesensaenderung der Epileptiker.* (Constitution and character changes in epileptics.) Leipzig: Georg Thieme, 1938.
11. Stotsky, B. A. A comparison of remitting and nonremitting schizophrenics on psychological tests. *J. abnorm. soc. Psychol.,* 1952, **47**, 489–496.
12. Weiszenfeld, F. P. Der Rorschachsche Formdeuteversuch als Hilfsmittel zur Differentialdiagnose zwischen genuiner Epilepsie und Uebererregbarkeitsepilepsie. (The Rorschach test as an aid in the differential diagnosis between genuine epilepsy and "overexcitement" epilepsy.) *Z. ges. Neurol. Psychiat.,* 1941, **171**, 321–336.
13. Zimmerman, F. T., Burgemeister, B., & Putnam, T. Intellectual and emotional makeup of the epileptic. *Arch. Neurol. Psychiat.,* 1951, **65**, 545–556.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

A Note on Some Projected Familial Attitudes Associated with Rorschach Movement Responses

Jerome L. Singer
Franklin D. Roosevelt VA Hospital, Montrose, New York

and Daniel A. Sugarman
Teachers College, Columbia University

This study represents a partial replication of aspects of an investigation by Shatin (5) which involved comparison of Rorschach and Thematic Apperception Test variables. Shatin reported that a significant association emerged between the number of Rorschach human movement responses (M) and the designation of parental figures on the TAT as "benign or nurturant." His subject sample consisted of 90 psychoneurotic patients. Recently, an increasing body of experimental data has tended to support Rorschach's empirical association of M responses with motor inhibition or delaying capacity, on the one hand, and fantasy or imaginativeness, on the other (2, 7, 8). Shatin's finding offers a clue to the origin of this capacity for delay presumably tapped by Rorschach's M. Since individuals who form numerous M responses tend also to describe parental figures as more benign, one can speculate that individual differences in fantasy or delaying capacity may have their origin in differing patterns of parent-child relationships. Perhaps the child who has at least some sense of security in his dependent relationship with one or both parents may be more capable of restraining motor response directed toward gratification of needs during the temporary absence of the parent. Such restraint might be accomplished by fantasied identification with the activity of the adult. In the case of the relatively consistent and benign parents the assurance of their return may reinforce the delaying capacity and the employment of fantasy as a temporary substitute. However tentative such a formulation, it seems worth while to attempt at least to confirm Shatin's finding using different subjects and another nosological group.

The principal hypothesis of the present study is that schizophrenic adults differing in production of Rorschach M differ predictably in their descriptions of parent-child relationships and parent identifications in Thematic Apperception Test stories. Specifically, it is hypothesized that patients designated as members of a High-M group attribute relatively more positive qualities to parental figures, identify adult figures more often as parents, and ascribe more positive features to the interpersonal relationship of parent and child than do Low-M patients.

Procedure

Both Rorschach and TAT protocols were obtained from 60 hospitalized veterans diagnosed as schizophrenics. The *S*s were divided into two groups, High-M and Low-M, using the production of two or more M as the cutoff point. The two groups did not differ significantly in other Rorschach scores nor in pertinent background factors, including age, IQ, diagnostic group, duration of illness, and educational experience. A further description of this sample was presented in connection with a previous study (7).

To insure relative objectivity all TAT protocols were analyzed in accordance with a scheme described in previous studies (3, 6). Only Cards 1, 2, 6BM, and 7BM of the TAT were included. The rating of the protocols was carried out by the authors independently without knowledge of *S*s' Rorschach group. Since scoring called only for judgments of the presence or absence of certain qualities, chi square was employed as a test of significance of the association between distributions of pertinent TAT ratings for the High- and Low-*M* groups.

Results

1. On Card 2BM ("Farm Scene"), High-*M* *S*s showed a slight trend in more frequently designating the right-hand female figure as "Mother." ($\chi^2 = 2.2$; $p/2 = .08$)

2. On Card 6BM ("Mother-Son"), Low-*M* *S*s more frequently failed to designate the female figure as "Mother." ($\chi^2 = 3.2$; $p/2 < .05 > .02$)

3. Both High- and Low-*M* *S*s did not differ in designating the "mother-figure" on Card 6BM as "cold or rebuking." Since 85% of the total sample so designated the "mother-figure," little discrimination could be expected here.

4. High-*M* *S*s described the atmosphere of the interpersonal relationship between "mother-figure" and son on Card 6BM more often as "warm" or possessing some positive elements and less often as "cold" than did the Low-*M* *S*s. ($\chi^2 = 5.9$; $p/2 < .01$)

5. High-*M* *S*s more often ascribed purely positive qualities to the "father-figure" on Card 7BM than did Low-*M* *S*s. ($\chi^2 = 4.4$; $p/2 < .03 > .01$)

6. High-*M* *S*s described the atmosphere of the father-son interpersonal relationship more often as "warm" or possessing positive elements and less often as "cold" or "neutral" than did Low-*M* *S*s. ($\chi^2 = 4.1$; $p/2 = .02$)

7. No differences emerged in ratings of "means-end cognizance" or in the nature of problems discussed on the four cards.

Discussion

The results in general seem to confirm the hypothesis and are in accord with Shatin's findings for a neurotic sample. Since it must be remembered that these schizophrenics present considerable discord in the interpersonal relationships depicted in their stories, the finding of differences on this dimension between High- and Low-*M* *S*s is all the more striking. As a matter of fact, these results are of interest because they tend also to offer with a different sample further support of conclusions reached in an earlier study comparing projected familial attitudes of schizophrenics and normals (6). Although social class status was not studied here (most *S*s in this group being of lower-middle or borderline status) the marked tendency for schizophrenics to ascribe more rejection and negative characteristics to parents is borne out again in this sample. Little change in percentages from those reported for the previous sample of 60 schizophrenics emerged.

The generally negative view of parental figures appearing in TAT protocols of the schizophrenics is modified in part in the records of patients with two or more Rorschach *M* responses. Pending further verification of this result with other samples and normal groups, it seems premature to draw any final conclusions. Nevertheless, an avenue of exploration into possible genetic factors involved in the *M* response and the general dimension of motor inhibition and fantasy seems to beckon. These findings suggest the possibility that some persons have developed long-standing positive fantasies of parental figures which sustain them during periods of delayed gratification. The results reported here are consistent with views expressed by Mowrer (4) and by various psychoanalytic theorists (1) that socialization or the evolution of the reality principle in ego development is related to incorporation of parental attitudes through positive identifications with adults.

Summary and Conclusions

This investigation was designed to test with a different sample and nosological group the finding reported by Shatin (5) that neurotic *S*s with relatively numerous Rorschach *M* tended to portray parental figures in TAT stories as "benign or nurturant." A sample of 60 hospitalized schizophrenic adult males was divided into High- and Low-*M* groups. Their protocols for TAT Cards 1, 2, 6BM, and 7BM were rated in categories relating to identification of parental figures and positive or negative characteristics ascribed to parents or to the atmosphere of the parent-child relationship. Results revealed that High-*M S*s tended more frequently to identify older female figures as mother and generally ascribed either somewhat more positive or less negative characteristics to fathers or to the interpersonal relationships between parents and children in these cards. The marked trend for these schizophrenic *S*s to perceive mother figures as cold or rejecting and to ascribe negative characteristics generally to parent-child relationships in their stories confirms an earlier finding with another sample of patients (6). Despite this general assignment of negative traits to parental figures, the emergence of somewhat more positive parental characteristics in the stories of High-*M S*s confirms Shatin's results. It is suggested that a fruitful approach to the understanding of the genesis and evolution of the Rorschach *M* and the capacity to delay overt response may involve exploration of the atmosphere of early parent-child relationships.

Received July 8, 1954.

References

1. Lair, W. S. Psychoanalytic theory of identification. Unpublished doctor's dissertation, Harvard Univer., 1949.
2. Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.*, 1953, **21**, 400–410.
3. Mitchell, H. E. Social class and race as factors affecting the role of the family in Thematic Apperception Test stories of males. Unpublished doctor's dissertation, Univer. of Pennsylvania, 1950.
4. Mowrer, O. H. Identification: a link between learning theory and psychotherapy. In *Learning theory and personality dynamics*. New York: Ronald, 1950. Pp. 573–616.
5. Shatin, L. Rorschach adjustment and the Thematic Apperception Test. *J. proj. Tech.*, 1953, **17**, 92–101.
6. Singer, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *J. consult. Psychol.*, 1954, **18**, 99–104.
7. Singer, J. L., & Herman, J. Motor and fantasy correlates of Rorschach's human movement responses. *J. consult. Psychol.*, 1954, **18**, 325–331.
8. Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult. Psychol.*, 1954, **18**, 1–9.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Relationship Between Diagnostic Category and Deviant Verbalizations on the Rorschach

William T. Powers and Roy M. Hamlin
Western Psychiatric Institute, University of Pittsburgh

In a study of pathological thinking in the Rorschach, Watkins and Stauffacher (6) compared 25 normals, 25 neurotics, and 25 schizophrenics on the basis of their "delta scores." The authors used a modification of Rapaport's types of deviant verbalizations on the Rorschach, which were felt by Rapaport to be more or less indicative of the presence of schizophrenia, and assigned quantitative values to various types of deviant responses. Watkins and Stauffacher obtained these values by having each author arbitrarily and independently assign, to each class of responses, a value which he felt to be appropriate on the basis of his experience. The specific values ranged from .25 for mildly pathological responses, through .50 for moderately pathological responses, to 1.00 for very pathological responses. Having developed this scoring procedure, they then proceeded to score, independently, the 75 Rorschach protocols used in their study. In order to equate for different length of protocols, they divided the total score for each protocol by the total number of responses in the protocol. In each case, this figure was then multiplied by 100 to obtain the delta score for the protocol. The average delta scores for the normal, neurotic, and psychotic groups were respectively 2, 4, and 18; and *t*-test comparisons significantly differentiated the normal group from the neurotic group and the neurotic group from the psychotic group at the .01 level of confidence. Watkins and Stauffacher report a scoring reliability coefficient of .78, highly significant, which is based on total scores for all 75 records.

The present paper reports results obtained with the method of scoring developed by Watkins and Stauffacher. The outside criterion in this case consisted of five clinical groups, two of them nonschizophrenic, and the other three schizoid or schizophrenic. A specific point of interest in planning the present study was the relative efficiency of Watkins and Stauffacher's scores as compared with scores of formal Rorschach elements, "sign" scores, and clinicians' judgments. Brief comment will be made on this point in the Discussion.

The hypothesis tested in this study was: Watkins and Stauffacher's scores show a significant correlation with the five clinical groups, representing degrees of psychological disorganization.

Method

Fifty subjects were used, 10 in each of five groups. The first group consisted of 10 socially adjusted subjects. Each of the other four groups may be labeled with a diagnosis. However, diagnostic label was never accepted as a sufficient criterion for inclusion in a group. For example, group 4 consisted of 10 patients with a diagnosis of paranoid schizophrenia. Each case record was studied carefully to insure that these 10 subjects met the following criteria: agreement on diagnosis over a reasonable period of time; definitely elicitable delusions; ideas of reference or hallucinations; a history of inadequate sexual adjustment; and a record of instances in which the subject indulged in bizarre behavior. The paranoid schizophrenic group was differentiated from the catatonic schizophrenic group by the absence of mutism, posturing, and other catatonic symptoms, with these factors in turn operating as criteria for inclusion in the catatonic schizophrenic group.

Each of the other four groups, including the normals, was screened with equal care and, of course, was selected for inclusion in the study before taking the Rorschach. This procedure insured somewhat more homogeneous and meaningful clinical categories than the uncritical use of diagnostic labels. A second purpose was to minimize contamination; a patient may be diagnosed as a paranoid schizophrenic partly because of his Rorschach. Every effort was made to include in the criteria for a group only acceptable clinical evidence and observation that could hardly come from the Rorschach itself.

The five groups were:

1. Socially adjusted subjects
2. Anxiety neurotics
3. "Latent" schizophrenics
4. Paranoid schizophrenics
5. Catatonic schizophrenics

Essentially, these five groups were selected to represent a rough continuum of "schizophrenic" disorganization, in line with Rapaport's description of deviant responses as characteristic of schizophrenia. However, a neurotic group was included, since Watkins and Stauffacher found that neurotic subjects obtained higher deviant response scores than normal subjects. The inclusion of a neurotic group does not imply any assumption of basic similarity between neurosis and schizophrenia. The hypothesis tested was that these deviant response scores are related to levels of psychological disorder or disorganization.

The catatonic schizophrenics were placed last in the list of five groups on the basis of consultation with psychiatrists, suggesting that they might show more disorganization in terms of deviant responses than paranoid schizophrenics.

The scoring used in this study followed Watkins and Stauffacher's procedure in all respects. In the present study, each Rorschach protocol was typed, and information as to age, sex, and education of the subject was given at the top of the protocol. In the scoring of each record, then, it seems apparent that the score assigned to any given response by a judge may be influenced by the opinion which the judge has concerning the previous responses which he has scored in this same record. In other words, each response was not evaluated *in vacuo* but in the context of all other responses in a given protocol since the scorer always had the total Rorschach before him. This may possibly be a weakness in the method of scoring, but does bring the procedure closer to a "global" approach.

Results

Validity. The mean delta scores which Watkins and Stauffacher obtained for their normal, neurotic, and psychotic groups were, respectively, 2, 4, and 18. In the present study, the mean delta scores for the normal, neurotic, latent schizophrenic, paranoid schizophrenic, and catatonic schizophrenic groups were, respectively, 6, 11, 21, 21, and 34. It is apparent that the mean delta scores in the present study are higher than those reported by Watkins and Stauffacher. The order of groups was exactly as predicted, although the difference between the latent schizophrenic and paranoid schizophrenic groups was minimal. An analysis of variance design was used and the F ratio was 6.2. An F ratio of 3.8 is significant at the .01 level for 4 and 45 degrees of freedom. An epsilon coefficient, or corrected eta, was also com-

Table 1

Number of Cases Falling in the Scoring Categories Indicated

Delta scores	Normals	Anxiety neurotics	Pre-schizophrenics	Paranoid schizophrenics	Catatonic schizophrenics
Above 50				1	1
45–49					
40–44			1	1	2
35–39			1		2
30–34					1
25–29			1		
20–24			1	2	
15–19		3	1	3	2
10–14	1	1	4	1	1
5–9	5	6	1	2	1
0–4	4				
Mean	5.62	11.40	20.53	21.00	33.50
SD	2.71	4.64	11.59	13.50	21.57

puted. This value was .65, a highly significant figure.

Table 1 shows the distribution of the subjects in the five groups with respect to their delta scores. The mean and standard deviation of each group are also indicated. Inspection of this table suggests cutoff points at scores 10 and 20. Only one subject, or 10%, of the normal group scored higher than 10, whereas 80% to 90% of each of the psychotic groups scored above 10. However, 40% of the neurotic group obtained scores above this cutoff point. A cutoff point at score 20 shows no normal or neurotic subjects obtaining higher scores, whereas 40% to 60% of each psychotic group scored above this point. Table 2 indicates the percentage of cases in each group falling above the selected cutoff scores.

Table 2

Percentage of Cases in Each Group Falling Above the Indicated Delta Scores

Delta score	Normals	Anxiety neurotics	Pre-schizophrenics	Paranoid schizophrenics	Catatonic schizophrenics
Above 20	0	0	40	40	60
Above 10	10	40	90	80	90

Any practical value that such cutoff points might have is, however, contingent upon strict consistency of scoring. The delta scores in the present study show a constant trend toward relatively higher values than those reported by Watkins and Stauffacher. This difference in scores is reflected in their suggested cutoff points at scores 5 and 10. Careful standard training of scorers is, of course, indicated, but there is also apparent need for further refinement and specification of scoring categories.

Scoring reliability. In Watkins and Stauffacher's study, the reliability of scoring coefficient, based on total scores, for all 75 records was .78, a highly significant figure. In the present study, only one author scored all 50 Rorschachs. Fifteen protocols were selected at random and scored by the second author to provide an estimate of reliability. Each scorer worked independently, after careful review of Watkins and Stauffacher's definitions and examples of the various scoring categories. A product-moment correlation coefficient was computed on the basis of the total scores assigned to each of these 15 Rorschachs. The obtained value of .88 supports the result obtained by Watkins and Stauffacher.

When the specific responses in each protocol are considered, however, the results are less satisfactory. The total number of responses in all 15 protocols amounted to 441. Of this total, one author scored 133 responses while the second scored 142 responses. The total number of jointly scored responses was 87, with the percentage of agreement slightly over 60. In the individual protocols the percentage of agreement ranged from 50 to 86.

The results are even less satisfactory when the categories to which the jointly scored responses were assigned are considered. Seldom was a jointly scored response assigned to exactly the same category by both judges. These results suggest that the scoring categories should be more adequately delimited, and those categories which are extremely unreliable or which serve little function in differentiating degrees of pathology should be eliminated or at least modified.

Watkins and Stauffacher have considered the problem of consistent scoring in some detail, and their data point to the same difficulties suggested here. One important indication from their results should be repeated: the reliability of scoring increases consistently with degree of psychological disorganization. Specifically, they report separate scoring reliabilities for their 25 normal, 25 neurotic, and 25 psychotic subjects. These coefficients are, respectively: .04, .47, and .91. Although the increasing reliability is related simply to the increasing magnitude of scores in the more pathological groups, two other factors should be considered. First, the "mildly deviant" responses, as specified at present, may be well nigh impossible to score with any suitable degree of reliability. Second, in both studies, the scorers worked with the entire protocol before them, not with one strictly isolated response at a time. The higher reliability for the more pathological groups may reflect global estimates of the Rorschachs, guided by specific attention to one response at a time. The tendency, or constant error, of the scorers in the present study to obtain higher values than those reported by Watkins and Stauffacher may be related to these considerations.

In spite of these scoring problems, it is nevertheless clear that encouraging coefficients of correlation were obtained for both reliability of scoring and for validity when total scores were used in conjunction with a correction factor for variability in number of responses, and when each independent worker scored all records considered in any given treatment.

Discussion

The specific focus of interest in undertaking the present study was the relation of Watkins and Stauffacher's method to other treatments of Rorschach data. The delta scores used here are superficially similar to Rorschach sign or determinant scores, or to judgments of small units of Rorschach material. Actually the delta scores represent a method midway between the sign approach and the use of global judgments. Hamlin (4) has previously reported on a series of 10 studies using variously the sign, small unit judgment, and global judgment approaches. These studies suggested that small, mechanically scored or judged units tend to give negative results when correlated against an outside criterion, and also that global judgments tend to give negative results when the material to be judged is too complex. The best results seem to be obtained in those situations where larger units than signs are used, and where samples of behavior sufficient to represent some degree of organization are used so that a clinical judgment is possible rather than a mechanical scoring without a rationale.

The method followed with the delta scores in the present study incorporates several features of interest in this respect. First, the scorer always worked with the total Rorschach before him. Second, his attention was directed to a small unit of projective behavior, a single Rorschach response, at a time. Each score or judgment was made on a small unit of material, but in the more global context of a total Rorschach. Third, the scoring involved an underlying rationale based on Rapaport's formulations. Thus, the scorer had a series of feasible tasks, each in a larger configurational context, and knew what he was doing in terms of a meaningful rationale.

The results suggest that these conditions contribute to an encouraging degree of correlation with an outside criterion of psychological disorganization. For example, as a rough indication, the obtained epsilon coefficient of .65 compares favorably with the values of .47 and .50 cited by Hamlin (4) from two studies (1, 3) where the outside criterion groups (representing degrees of psychopathology) and the variable judged (overall adjustment) were somewhat similar. The correlation of .68 cited from another study (2), where Wechsler-Bellevue IQ was the outside criterion, is not significantly higher.

The general questions underlying this and the other studies mentioned have been: under what conditions can clinicians make valid judgments from projective material? and how do successful judges proceed in making predictions? The possible practical value of such scales as Watkins and Stauffacher's for routine clinical use is questionable. Techniques like the Rorschach will continue to be used as clinical tools without being reduced to an intricate variety of objectively scored scales. The study of judgments and of conditions necessary for successful judgments may, however, clarify the efficiency of projective procedures.

Summary

The hypothesis tested in this study, that Watkins and Stauffacher's delta scores will show a significant correlation with the five diagnostic groups, was upheld. The *F* ratio of 6.2 and the epsilon value of .65 were both significant at the .01 level, as was also the .88 scoring reliability coefficient obtained from a comparison of total scores assigned by the two authors to 15 jointly scored protocols.

The significant results were in keeping with the results obtained by Cummings (3), Albee and Hamlin (1), and Watkins and Stauffacher (6) in similar studies, and the methods used in the present study resembled the methods used in those studies.

In all of these studies, as opposed to less successful judgment studies, the scorer or judge dealt with small units of material and made judgments based on a specific rationale. The procedure of allowing the judges to have the complete Rorschach protocols before them while evaluating the specific responses gives the added advantage of allowing the judge to gain an over-all impression of the record, while at the same time his forced consideration of each individual response insures that his final judgment of the protocol will by no means issue solely from his opinion of the total gestalt.

Received August 5, 1954.

References

1. Albee, G. W., & Hamlin, R. M. Judgment of adjustment from drawings: the applicability of rating scale methods. *J. clin. Psychol.*, 1950, **6**, 363–365.
2. Bialick, I., & Hamlin, R. M. The clinician as judge: details of procedure in judging projective material. *J. consult. Psychol.*, 1954, **18**, 239–242.
3. Cummings, S. T. The clinician as judge: judgments of adjustment inferred from Rorschach single-card performance. *J. consult. Psychol.*, 1954, **18**, 243–247.
4. Hamlin, R. M. The clinician as judge: implications of a series of studies. *J. consult. Psychol.*, 1954, **18**, 233–238.
5. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing*. Vol. 2. Chicago: Year Book Publishers, 1946.
6. Watkins, J. G., & Stauffacher, J. C. An index of pathological thinking in the Rorschach. *J. proj. Tech.*, 1952, **16**, 276–286.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Measurement of Anxiety in Clinical Situations[1]

Arnold H. Buss, Morton Wiener, Ann Durkee, and Marc Baer
Carter Memorial Hospital, Indianapolis

Anxiety has been measured in a variety of ways, but there is little general agreement among measures. A possible source of this disagreement is the methodological problems involved in evaluating anxiety. The problems encountered in measuring clinical anxiety are primarily those of definition and reliability of measurement. Because the term "anxiety" has a variety of behavioral referents, the operational definition should be explicit, and the behaviors labelled "anxious" should be specified. When anxious behavior is rated in clinical situations, there should be several judges and a statement of interjudge agreement.

Two previous studies illustrate attempts at solving these problems. Gleser and Ulett attempted to measure "anxiety-proneness," which was defined as "the present mental status, past history of anxiety, background and personality development that would indicate the occurrence of anxiety symptoms in a stressful situation" (4, p. 280). A psychiatrist and a psychologist used an eight-point scale to rate anxiety-proneness. The interrater reliabilities were .51 for patients and .30 for normals. Elizur defined anxiety as "an inner state of insecurity which may take one or more of the following forms: fears, phobias, lack of self-confidence, extreme shyness, ideas of reference, and marked sensitivity" (3, p. 248). From protocols of interviews with college students three judges made anxiety ratings on a nine-point scale. The average interrater correlation was .70.

In evaluating these two studies it may be seen that Elizur's definition of anxiety is more specific than that of Gleser and Ulett. This was evidently reflected in a considerably higher interjudge reliability in rating anxiety in normals: .70 for Elizur and .30 for Gleser and Ulett. It seems likely that further specificity in defining anxious behaviors would yield still higher interjudge reliability and would prove more useful in future research in this area. If reliability can be established, the obtained measure could be used as a criterion for validating other measures, such as anxiety questionnaires.

Purpose

The purposes of this paper were: (*a*) To attempt a clear delineation of various manifestations of anxiety and to determine if this results in higher interrater reliability; (*b*) to determine the relationship of each manifestation of anxiety to a global evaluation of anxiety; (*c*) to use the ratings of anxiety as a criterion against which to validate the Taylor Scale of Manifest Anxiety (*A* scale) (7).

Method

Subjects. The *S*s were patients in a neuropsychiatric hospital. Patients were excluded from the study for the following reasons: (*a*) out of contact, (*b*) on electroshock or insulin therapy, (*c*) low intelligence, (*d*) diagnosed organic brain damage. Seventy-three patients (33 men, 40 women) were interviewed, and 64 of these completed the *A* scale. Of the nine *S*s who failed to complete the *A* scale, several stopped in the middle and refused to go on, and the remainder complained of being ill.

Procedure. Each *S* was interviewed in a room with four psychologists, who served as judges. A pilot study had indicated that the presence of additional persons in the interviewing room did not markedly affect *S*'s anxiety level. Patients in this hospital are frequently seen by several members of the staff simultaneously. Each *S* was interviewed by the same psychologist, although the other three psychologists also asked questions at the end of the interview. The interview lasted approximately 15 minutes. Then the *S* was conducted to another room, and the *A* scale was administered by another staff member. Immediately after the interview, each judge rated *S* for anxiety. Each rating was made independently and without consultation with the other judges.

---

[1] This study was carried out with the permission and encouragement of Juul Nielson, M.D. The writers are indebted to Drs. Charles Windle and Leonard Goodstein for their critical reading of the manuscript.

The interview. The interview was relatively uniform. The interviewer introduced himself and then the other three judges to *S*. At first all questions were open-ended and unstructured, e.g., "What has been bothering you?" "What happened?" "How do you feel inside?" As the interview progressed, the questions became more specific and direct, e.g., "Do you have aches and pains?" "Do you have trouble sleeping?" "Does your heart beat fast?" At the end of the interview, the interviewer asked the other judges if they had any questions to ask. Finally, *S* was given the task of subtracting sevens from 100, and the interview was ended.

The ratings. Although definitions of anxiety in psychiatric and abnormal psychology texts (1, 2, 5, 6, 8) vary, there tends to be general agreement on the manifestations of anxiety. In this study an attempt was made to use only those manifestations of anxiety on which there is general agreement. It was decided to group symptoms into categories because the probability of occurrence of any given symptom, e.g., excessive sweating, in any given *S* was small. Each category was rated on a five-point scale: a rating of 1 represented absence of or minimal manifestation, and a rating of 5 represented maximal manifestation.

The categories were divided into behavior observed and behavior reported by the patient in the situation, as follows:

A. *Observed*

1. *Distractibility*. It was found in the pilot study that distractibility per se was difficult to rate. Therefore a task was introduced, the task of subtracting sevens from one hundred. Serial subtraction is typically used in psychiatric settings as a measure of distractibility. Inattentiveness during the interview was also taken into account in rating this category.

2. *Restlessness*. This category involved the skeletal components of anxiety: agitation, tremors, and tics. It was arbitrarily decided to rate at least 4 (on the five-point scale) for the presence of tremor and to rate 5 for marked tremor.

3. *Physiological concomitants*. This category involved the autonomic components of anxiety. Four manifestations were selected: sweating, flushing, breathing disturbance, and excessive swallowing. The rating was based on both intensity and number of symptoms.

B. *Reported*

4. *Subjective feeling of tenseness*. This category referred to the patient's report of how much he was bothered or upset. The rating was based on the intensity, frequency, and chronicity of the complaint. It was arbitrarily decided that the verbalization of panic would be rated 5.

5. *Worry*. This category referred to apprehension, regardless of whether there was a basis in fact. It was arbitrarily decided to rate 2 if *S*'s worry concerned only money or leaving the hospital. The rating was based on intensity and chronicity.

6. *Somatic complaints*. This category referred to aches, pains, and peculiar feelings exclusive of actual illness, e.g., gallstones, arthritis, or heart disease, although somatic delusions were included. The ratings were based on the number, frequency, intensity, and chronicity of the complaints.

7. *Physiological concomitants*. This category involved the same manifestations as observed physiological concomitants, with the addition of palpitation. It might be expected that the observable concomitants would be the same as the reported concomitants. However, it was demonstrated in the pilot study that these two categories were not identical; several patients reported physiological concomitants without manifesting them in the interview situation.

8. *Muscular tension*. This category involved the same manifestations as observed restlessness, with the addition of a feeling of internal tension. At first glance, muscular tension may appear to be identical with subjective feelings of tenseness. However, as defined in this study, it referred to an awareness of muscular contraction that was reported as "tied up in knots," "ready to explode," "feel shaky." This may be contrasted with the reports on subjective feelings of tenseness: "nervous," "upset," "stirred up inside," "panicky."

9. *Over-all rating of anxiety*. This category refers to a global estimate of anxiety. The ratings were based on what may be termed a clinical impression of anxiety. The judges attempted to divorce this rating from any particular category or configuration of categories.

Results

The anxiety ratings of each judge were compared with those of every other judge. Thus, for each of the nine categories there were six correlations.[2] The mean correlations between judges are shown in the first two columns of Table 1. The mean correlation of .83 between judges for over-all anxiety indicates fairly high interrater reliability.

The ratings for distractibility appear somewhat more reliable than the ratings in the other categories. The rating for distractibility was based largely on the patient's performance on the task of subtracting sevens. The number of errors and the amount of time taken to perform this task are fairly objective measures, which probably accounts for the high reliability.

The ratings for observed physiological concomitants seem somewhat less reliable than the ratings of the other categories. This may be ascribed to a fundamental difficulty in observing autonomic changes. Changes in the depth and rate of breathing are discernible, but changes in sweating, flushing, and swallowing are often so minimal as to escape notice. The ratings for the remaining categories yielded mean intercorrelations that ranged from .66 to .85, indicating a moderate degree of reliability in assessing these aspects of anxiety.

The eight individual categories of manifest anxiety were selected on the basis of their expected relationships to global anxiety. In order to determine these relationships the ratings in each category were pooled [3] and then compared with the pooled ratings of over-all anxiety. The correlations, shown in the third column of Table 1, indicate that there is a definite relationship between the over-all rating of anxiety and the ratings in each individual category. While the correlations range from .47 to .72, there is little difference between adjacent correlations, and the differences may be ascribed to chance fluctuation.

Table 1

Intercorrelations Among Judges, and the Correlations of the Ratings with Over-all Anxiety and with the Taylor *A* Scale

Category	Intercorrelation among judges: Mean	Range	Correlation with over-all rating	Correlation with *A* scale
Observed				
1. Distractibility	.93	.92–.94	.47	.16
2. Restlessness	.74	.68–.80	.66	.37
3. Physiological concomitants	.56	.41–.65	.56	.22
Reported				
4. Subjective feelings of tenseness	.66	.59–.77	.65	.52
5. Worry	.76	.71–.79	.50	.50
6. Somatic complaints	.85	.81–.88	.52	.40
7. Physiological concomitants	.81	.76–.86	.72	.68
8. Muscular tension	.73	.67–.77	.68	.54
9. Over-all rating of anxiety	.83	.66–.92		.60

There were two categories of physiological concomitants, one observed and one reported. It might be expected that there would be a "halo-effect," resulting in a high degree of relationship between these categories. However, the correlation is .30, indicating a rather low degree of relationship. Evidently, *S*'s report of autonomic disturbance differed from what the judges observed in the interview situation.

Similarly, restlessness and muscular tension referred to essentially the same manifestations of anxiety, the former observed and the latter reported. The correlation between ratings in these categories is .61, indicating a definite relationship. Evidently there is a greater correspondence between observed and reported skeletal symptoms than there is between observed and reported autonomic symptoms.

Next, the pooled ratings for each category and for the over-all rating of anxiety were compared with the Taylor *A*-scale scores. The correlations are shown in the last column of Table 1. There is a low relationship between the *A* scale and the observed categories and a moderate relationship between the *A* scale and the reported categories. The correlation for reported physiological concomitants was somewhat higher than that for the other categories, which suggests that there are many items of this sort in the *A* scale. The over-all rating of anxiety correlated .60 with the *A* scale, which suggests that the *A* scale measures many of the same kinds of behavior as the ratings do.

---

[2] All correlations in this study are Pearson product-moment *r*'s; the correlations were averaged by the *z* transformation method.

[3] It seems reasonable to assume that the best estimate of each *S*'s rating for any given category is the consensus of the four judges. Therefore the ratings of the judges were averaged, and these pooled ratings were used in all subsequent comparisons.

The over-all rating of anxiety was a "clinical" integration of the eight categories of anxious behavior. Integration may also be obtained by arithmetic summation of the ratings in each category; the pooled ratings for categories 1 to 8 were added to yield a single score for each *S*. The correlation between these arithmetic sums and the over-all ratings is .93, which suggests that there is little difference between arithmetic and clinical integration. The arithmetic sums of each judge were correlated with those of every other judge. The mean intercorrelation is .88, and the range is .81 to .91. The correlation between the pooled arithmetic sums and *A*-scale scores is .64. This is not significantly different from the correlation of .60 between the over-all ratings and *A*-scale scores.

Discussion

In this study clear delineation of the manifestations of anxiety resulted in reliable ratings. The average intercorrelation between judges of .83 may be compared to the correlation of .51 obtained by Gleser and Ulett (4) on their ratings of patients. With normal subjects Elizur (3) obtained an average interjudge correlation of .70, while that of Gleser and Ulett was .30. Both Elizur's definition and the present one were more specific than that of Gleser and Ulett. Thus, it may be concluded that specificity of definition increases interjudge reliability in the measurement of anxiety.

It has been assumed heretofore that the various aspects of anxiety must be integrated more or less intuitively in order to obtain a global measure of anxiety. This procedure was followed in the present study, in which an over-all clinical rating of anxiety was made by each psychologist. A likely basis for assuming the necessity of such intuitive integration is that there must be an implicit "weighting" of the various aspects of anxiety. The psychologist is expected to use his clinical judgment in deciding which aspects are more important in any given *S*. The question arises as to the advantages of clinical integration over mere arithmetic summation. The results showed an extremely high relationship between these two types of integration and that arithmetic summation yields slightly higher interjudge reliability. Apparently, clinical integration offers no advantages over arithmetic summation. Thus, it may be concluded that specific denotation of the various manifestations of anxiety, followed by summation of the ratings, yields an adequate, reliable measure of anxiety. It would appear that there is no necessity for vagueness in defining or measuring anxiety.

The manifestations of anxiety listed in most texts include both observed and reported behavior. The definition of anxiety in this study included both aspects. This is reflected in the relationship between each type of behavior and the over-all clinical rating. The correlations between the ratings for the observed categories and the over-all clinical rating were of approximately the same order as the correlations between the ratings for the reported categories and the over-all rating. Thus the over-all rating did take into account both types of behavior.

Since the clinical evaluation of anxiety conformed to a consensual definition of anxiety and proved to be reliable, it was used as a criterion measure. The *A* scale showed a moderately high relationship to both the arithmetic summation and the over-all clinical estimate of anxiety. The *A* scale showed a much higher relationship to the reported categories than to the observed categories. This was expected because the *A* scale consists only of self-ratings, i.e., reported manifestations of anxiety. Anxious behaviors observed by others are omitted. To the extent that the *A* scale omits items that relate to observed measures of anxiety, it does not completely measure anxiety, as defined by psychiatric and abnormal psychology texts. However, the *A* scale's relationship to the criterion suggests that it is a fairly adequate measure of anxiety.

It should be possible to construct an anxiety inventory with an extremely high relationship to an external criterion. This might be done by performing an item analysis on the *A* scale, using the clinical ratings to select high-anxious and low-anxious subjects. Such an inventory would provide a quantitative estimate of anxiety, as defined clinically. The advantage over present-day anxiety inventories, which have only face validity, is obvious.

Summary

This study attempted to evaluate anxiety (as defined by textbooks) in a clinical situation and relate this measure to anxiety as measured by the Taylor Scale of Manifest Anxiety. Seventy-three neuropsychiatric patients were interviewed, and 64 of these completed the *A* scale. Four judges rated the *S*s for various aspects of anxiety on the basis of the interview material. The results were as follows:

1. Specific denotation of the manifestations of anxiety resulted in reliable ratings.
2. "Clinical" integration of anxious behavior to yield a global measure of anxiety correlated .93 with mere arithmetic summation of the various manifestations.
3. The correlation between the over-all rating of anxiety and *A*-scale scores was .60.

Received July 29, 1954.

References

1. Cameron, N. *The psychology of behavior disorders*. Boston: Houghton Mifflin, 1947.
2. Cattell, R. B. *Personality*. New York: McGraw-Hill, 1950.
3. Elizur, A. A content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, **13**, 247–283.
4. Gleser, Goldine, & Ulett, G. The Saslow Screening Test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8**, 279–283.
5. Henderson, D., & Gillespie, R. D. *A text-book of psychiatry*. London: Oxford Univer. Press, 1950.
6. Noyes, A. P. *Modern clinical psychiatry*. Philadelphia: Saunders, 1953.
7. Taylor, Janet A. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **41**, 285–290.
8. Weiss, E., & English, O. S. *Psychosomatic medicine*. Philadelphia: Saunders, 1949.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Cross Validation of the Organic Brain Pathology Scale for a Test of Symbol Arrangement[1]

Theodore C. Kahn
USAF School of Aviation Medicine, Randolph Air Force Base, Texas

The validation of the organic brain pathology scale of a technique which requires the subject to arrange 15 plastic objects several times on a segmented strip appeared in 1951 (1). The objects used in the test vary in color, size, and thickness, and their shapes represent easily recognized symbols which are common in our culture. The administration takes approximately 15 minutes and the scoring a like amount of time. In the original study 50 patients with a diagnosis of organic brain pathology were matched with a non–brain-damaged group. In that study 197 variables were statistically evaluated and 99 of these were found to be significant. The difference of the means yielded a *t* ratio of 17.6 when the distributions of the scores of the two groups were compared.

Thirty patients in a state mental hospital who had a diagnosis of organic brain pathology and whose medical histories roughly corresponded to the previous patient group were matched in age, IQ, educational level, race, sex, religion, and national origin with a non–brain-damaged group of like size. The screening of the control group was carried on under the supervision of a psychiatrist. Factors which were not equated and which were significantly different in the two groups were the number of subjects who had been veterans, the number who lived in rural versus urban environment, occupational level, and marital status.

All of the scores of the new organic group fell within the range of scores previously established by the first organic group with the exception that one patient scored lower (away from normal) than had any of the organics in the previous testing. Likewise, the scores of the new non–brain-damaged group fell within the range of scores of the first non–brain-damaged group except that the score of one subject, who was over 75 years old, fell somewhat outside of this range into the direction of the organic group. The mean score of the new organic group was -226.2 with a σ of 110.9; that of the non–brain-damaged control group was $+34.6$ with a σ of 65.2. The difference of the means of the scores of the two groups yielded a *t* ratio of 11.1. Since the analysis of variance produced a significant *F* ratio, the medians were compared by chi square and the difference was found to be significant at better than the .001 level.

Cutoff scores were established for the four groups in the two studies. If three subjects over 70 years old are excluded from the two non–brain-damaged groups, a cutoff score of -54 would separate 92 per cent of the first organic group, and 100 per cent of the second organic group, from all the non–brain-damaged subjects in both groups. Even when these three low scoring subjects are included, a cutoff score of -138 separates all the non–brain-damaged subjects in the two studies from 72 per cent of the first organic group and from 83 per cent of the second organic group.

Brief Report
Received January 12, 1955.

Reference

1. Kahn, T. C. An original test of symbol arrangement validated on organic psychotics. *J. consult. Psychol.*, 1951, **15**, 439–444.

---

[1] An extended report of this study may be obtained without charge from Major T. C. Kahn, USAF School of Aviation Medicine, Randolph Field, Texas, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4482 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Digit Symbol Performance as a Function of Increasing Levels of Anxiety

Joseph D. Matarazzo and Jeanne S. Phillips

Division of Psychosomatic Medicine and Department of Neuropsychiatry, Washington University School of Medicine

In two earlier studies (2, 3) we reported a significant curvilinear relationship between performance in a timed learning situation and anxiety level as measured by the Taylor Scale of Manifest Anxiety (*A* scale) (6, 7), and a significant rectilinear relationship between anxiety level and intelligence as measured by a timed intelligence test. The present study was designed to extend these findings by employing a timed learning task, the Digit Symbol substitution test, which in the past has been employed both as a measure of learning and as an intelligence subtest. With the traditional procedure the 67-item Digit Symbol subtest of the Wechsler-Bellevue (10) was extended to 175 items. These were printed on one side of a sheet of paper. The time limit was increased from 1½ minutes to 3 minutes. The directions used were the standard ones.

In a group administration, a total of 119 *Ss* was used in the experiment. This number included an entire class of first-year medical students ($N = 81$) and approximately one-half of the sophomore class ($N = 38$). It was learned before the experiment that one student had had previous experience with the digit symbol task and he was excluded from the study. The *A* scale was administered to all *Ss* immediately following the learning task.

The hypothesis to be investigated was that digit symbol performance would have a curvilinear relationship to increasing levels of anxiety. On the basis of the results of the earlier stylus maze study (3), we predicted a curvilinear relationship based on the following criteria: (*a*) a U-shaped curve as determined by visual inspection, (*b*) a higher value of epsilon than *r* and a correspondingly significant value for the chi-square test of linearity of regression, and (*c*) a significant difference between the means of groups 1 and 2, 1 and 3, 1 and 4, 1 and 5, 3 and 6, and 3 and 7. These differences in means for specific groups (criterion *c*) were predicted on the basis of the significant differences obtained for the time criterion in the original study. It was realized at the outset that these predictions were contingent upon the way in which *Ss* of the present study distributed themselves in groups according to their Taylor raw scores. If an insufficient number of *Ss* fell in any group, predictions about that group would necessarily no longer be valid.

Results

On the basis of their Taylor raw scores the 119 *Ss* were assigned, as in our previous studies, to groups on a scale of increasing anxiety level. Individuals with raw scores of 1–5 were given an anxiety rating of 1, while those with raw scores of 6–10, 11–15, 16–20, etc., were given ratings of 2, 3, 4, etc., respectively. While the over-all distribution of Taylor raw scores for *Ss* of this study (ranging from 1–35, mean of 12) was similar to that reported by Taylor (7) and also to that found with the maze and intelligence data, the number of *Ss* in each subgroup (especially the groups at the high end) differed slightly from our previous findings. Thus, only 11 *Ss* earned Taylor raw scores above 21 (five from 21–25, four from 26–30, and two from 31–35). In an effort to avoid the problems imposed by these small *N*'s, the 11 *Ss* were grouped together. This left five groups: group 1 included raw scores 1–5, and individuals with raw scores of 6–10, 11–15, 16–20, and 21–35 were placed in anxiety groups 2, 3, 4, and 5. The number of *S*s in each of these five groups from low to high was 18, 41, 34, 15, and 11, respectively.

Before the mean digit symbol performance of each of these five groups was computed, criterion *c* of the original prediction was changed to conform to these findings. The prediction was now changed to read: based on the results of the maze study, the following groups will be expected to differ significantly: 1–2, 1–3, 1–4, and 1–5. Although groups 3–5 did not differ significantly in the earlier study (groups 3–6 and groups 3–7 did), it is possible that 3–5 will differ in this study since group 5 now includes *S*s with raw scores ranging from 21–35, and thus overlaps groups 5, 6, and 7 of the previous study.

The mean digit symbol scores for each of the five groups are shown in Fig. 1. The values from low to high anxiety are: 101.8, 109.7, 109.8, 108.8, and 106.4. These values and visual inspection of the curve confirm criterion *a* of our hypothesis, i.e., that the function is a U-shaped curve. Two aspects of the curve are worth noting: (*a*) as with the time measure of the maze data, the largest mean difference is found between groups 1 and 2, and (*b*) the failure of the curve to make a complete U at the high end. This latter may be a function of the unfortunate restriction in the number of groups at the high end.

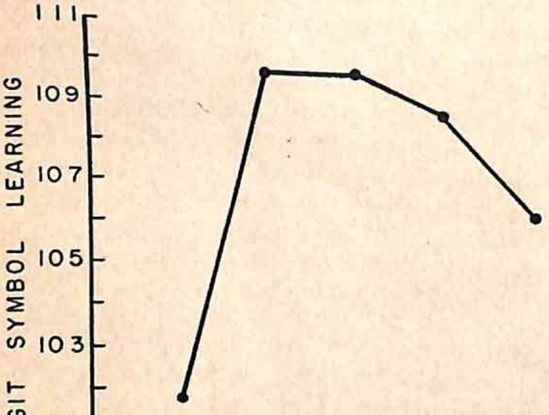

Fig. 1. Mean digit symbol score for each of the five groups ranging from low anxiety (group 1) to high anxiety (group 5).

Since the data in Fig. 1 represent only the means of the learning scores in each of the five groups, the confirmation of the hypothesis, by visual inspection alone, is not possible. This is especially true since there was considerable variability in learning score from individual to individual within each of the five groups (the sigma was roughly of the order of 15 for each of the five groups).

Criteria *b* and *c* were next applied. As in the two previous studies (2, 3) the *F*-test value of 1.02 was not significant. The reasons for this are discussed in the earlier studies and involve the fact that there may be a *functional* relationship between the two variables (digit symbol performance and anxiety level) and that the five anxiety groups are not "independent" but ordered in a sequential relationship to each other. This relationship may increase the "within" variance while not influencing the "between" variance. This whole problem, as well as the permissibility of using *t* tests with data of this type, in the absence of a significant *F*, has been discussed by Webb and Lemmon (8, 9).

Analyses of the differences between the means for the various groups revealed that two of the four differences predicted were statistically demonstrated: groups 1–2 and 1–3 were each found to differ significantly at the .03 level of confidence. The remaining two differences, between groups 1–4 and 1–5, were not demonstrated. The difference between groups 1–4 was just short of the .05 level, however.

The additional prediction, that groups 3–5 might possibly differ significantly, did not hold. Despite the relatively moderate mean difference between these two groups, shown in Fig. 1, the intragroup variability was such that the difference was not statistically significant. This finding is unfortunate for the hypothesis since it was the only difference (due to the restricted number of *S*s at the high end) predicted between the high and middle ranges of the anxiety continuum. Consequently, criterion *c* for the establishment of a curvilinear relationship between digit symbol performance and increasing levels of anxiety was not confirmed. That it would have been confirmed if we had had two additional groups at the high end is a matter simply for conjecture. The demonstrated superiority of *S*s in the middle range of anxiety, groups 2 and 3, relative to *S*s in the low anxiety group, group 1, is nevertheless a significant finding.[1]

As was pointed out in the two previous studies, the most meaningful determination of the relationship between the two variables under discussion is the use of correlational techniques (criterion *b*). These methods allow determination of the *strength* of the relationship shown in Fig. 1. Both the correlation ratio, eta (1, pp. 259–261), and the unbiased correlation ratio, epsilon (4, pp. 319–324), can be computed from the analysis of variance data.

The values of both of these measures (.20 and .00, respectively) did not reach statistical significance and therefore do not differ from zero. This finding, which is unlike the data from the maze study, is surprising in view of the similarity between our present Fig. 1 and the earlier curve relating time to learn the maze with anxiety level. The value of the Pearson *r* was found to be an insignificant .05. Since neither correlation ratio is statistically significant, it was not possible to apply the chi-square test of linearity of regression (4, p. 319) in order to determine whether or not the true function deviates from a straight line.

Discussion

One is thus forced to conclude, within the limits of these data, that the hypothesis of a curvilinear relationship between digit symbol learning and increasing levels of anxiety—with optimal learning in the middle range—is not borne out. The hypothesis is supported to the extent that the relationship between the low anxious and middle anxious groups was demonstrated, but the equally important relationship (for the hypothesis) between the middle group and the high anxious group, despite the trend in means, was not borne out. The failure to demonstrate with the present task even the low-order, but nevertheless significant, correlations attained in the earlier studies, in view of the mean differences shown in Fig. 1, poses a question regarding the "meaning" of these same differences and also of the numerous similar ones reported in other studies using the *A* scale.

That there are, in fact, true differences in conditioning and learning abilities of *S*s who differ in Taylor anxiety score has been demonstrated many times.[2] The present study adds to this growing literature. However, the equally important finding of low correlations (when these have been computed) between anxiety level and performance in various situations implies what has long been known, that factors other than anxiety, or intelligence for that matter, are involved in learning ability. Accordingly, correlational techniques would appear necessary in the design of future studies utilizing the Taylor scale.

Summary

On the basis of previous work, a curvilinear relationship was predicted between Digit Symbol performance and score on the Taylor Scale of Manifest Anxiety. A total of 119 *S*s was divided into five groups ranging from low to high anxiety. The prediction was borne out in part, the performance of the middle groups being superior to that of *S*s in the low anxious group but, despite a trend in the predicted direction, the hypothesis was not confirmed for the middle and high anxious groups. The *strength* of the relationship, however, was found by correlational techniques to be an insignificant one. A methodological question regarding the use of the Taylor scale was raised, i.e., the need to include correlational methods of analysis in future studies because of the low-order relationships obtained in studies to date.

Received July 21, 1954.

References

1. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.

---

[1] Consistent with the finding of our previous study, the *S*s comprising the five groups of this study showed no differences in mean intelligence score as determined by the CVS abbreviated intelligence scale.

[2] It is of interest in view of the hypothesis of our present study that Spence and Taylor have recently reported (5, p. 270) finding a *curvilinear* relationship between conditioning and anxiety level in two unpublished studies.

2. Matarazzo, J. D., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, **18**, 201–205.
3. Matarazzo, J. D., Ulett, G. A., & Saslow, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, in press.
4. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
5. Spence, K. W., & Taylor, Janet A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.
6. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
7. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
8. Webb, W. B., & Lemmon, V. W. A qualification in the use of analysis of variance. *Psychol. Bull.*, 1950, **47**, 130–136.
9. Webb, W. B., & Lemmon, V. W. A sequel to the notes of Patterson and Diamond. *Psychol. Bull.*, 1952, **49**, 155.
10. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Discriminative Power of the Strauss-Lehtinen Battery in Adult Mental Patients[1]

A. A. Rafi

Netherne Hospital, Surrey, England

So far no experiments have been reported with adult subjects that duplicate the researches of Strauss and Lehtinen (1) with children. These authors emphasize the capacity of their battery to discriminate normals, endogenous mental defectives, and exogenous mental defectives but offer no quantitative evidence.

The present investigation aimed at determining whether the battery discriminated to a statistically significant extent adults of analogous categories, i.e., bright and dull mental patients without brain damage, mental defectives, and brain-damaged persons. Two statistical techniques were employed: analysis of variance by double classification to establish the general effect of diagnostic and sex differences on the separate tests of the battery, and discriminant function analysis to ascertain the efficiency with which the whole battery distinguished the particular categories.

Method

The test material, instructions, and testing procedures were modeled on those described by Strauss and Lehtinen (1). The Marble-Board, Figure-Background, and Tactual-Motor tests were quantified in the following manner:

Marble-Board Test (six patterns—maximum score 18)

Score

3 constructive performance of pattern;
2 orderly continuous sequence along outline of pattern;
1 discontinuous, sporadic performance with outline similar to pattern;
0 discontinuous, incoherent, sporadic performance with outline dissimilar to pattern.

Figure-Background Test (nine items—maximum score 27)

Score

3 correct naming of object with no reference to background;
2 noting presence of object in vague, imprecise manner as "round thing," "long thing";
1 mentioning both object and background but imprecisely, e.g., "scribbling with something round in the middle";
0 describing background, not mentioning object.

Tactual-Motor Test (three items—maximum score nine)

Score

3 figure only;
2 figure with contour of board;
1 figure and background;
0 background only.

The behaviors scored 0 are stated to be characteristic of brain-injured children (1).

The Dot-Perseveration Test was also administered, but perseveration appeared too rarely for adequate quantification in measurement terms and when treated enumeratively the differences were not statistically significant. This test was dropped from the battery.

All the subjects were inpatients at Netherne Hospital. Each sample of the categories studied comprised 20 subjects, 10 males and 10 females with age distributions as follows:

---

[1] Thanks are due to Drs. R. K. Freudenberg, physician superintendent, Netherne Hospital, for facilities given to carry out this research, and J. P. S. Robertson, senior psychologist there, for his encouragement and guidance.

Thanks are also extended to Drs. Doreen Firmin, physician superintendent, St. Lawrence's Hospital, and A. Walk, physician superintendent, Cane Hill Hospital, for additional cases to make up the cross-validating samples.

bright mental patients, mean 47.7, *SD* 9.47; dull mental patients, mean 45.7, *SD* 11.23; certified mental defectives, mean 34.2, *SD* 12.03; and brain-damaged patients, mean 50.2, *SD* 11.78. The first two categories were described psychiatrically as well-behaved chronic patients on parole. The allocation into bright or dull was made on the basis of Wechsler vocabulary raw scores, the limits of the bright category being 27.5 to the maximum, those of the dull category 18.5 to 22. None of the certified mental defectives had a Wechsler vocabulary raw score above 15. The neurological classification of the brain-damaged sample was as follows: paresis, five males and two females; cerebral atrophy, three males and four females; secondary cerebral carcinoma, one male; Huntington's chorea, four females; head injury, one male.

When the test results had been scored by the writer, they were independently scored by another person (not a psychologist) who was provided only with the actual results and the scoring instructions. There was only a very small proportion of disagreements; the product-moment correlation exceeded .90 for each test, using the difference method.

Results

Table 1 summarizes the results of the analyses of variance. It is clear that there are statistically significant differences among the categories in performance on each test. The factor of sex influenced only one test and that to a limited extent. The effect of the interaction was negligible.

In the case of the discriminant function analysis six equations were worked out: (*a*) bright versus dull; (*b*) bright versus mental defective; (*c*) bright versus brain-damaged; (*d*) dull versus mental defective; (*e*) dull versus brain-damaged; and (*f*) mental defective versus brain-damaged. These comparisons are set out in Table 2. The first column lists the comparisons as just enumerated. The second column gives the equations; here X = Marble-Board score, Y = Figure-Background score, and Z = Tactual-Motor score. The third column gives typical D for (A), the first member of a given comparison, and the fourth column typical D for (B), the second member of that comparison. The statistical significance of results is indicated as usual. In order to display the actual effectiveness of discrimination the cumulative percentages cut off by different values of D are shown in Table 3 for the four statistically significant comparisons and for the last comparison because of its special interest.

Table 1

Effects of Category and Sex: Analyses of Variance

		Mean squares		
Source	*df*	Marble Board	Figure Back-ground	Tactual Motor
Category	3	131.950***	428.245***	44.292**
Sex	1	18.050*	103.512	0.013
Interaction	3	7.017	48.213	5.266
Error	72	4.580	51.820	10.587
Total	79			

* Significant at .05 level.
** Significant at .01 level.
*** Significant at .001 level.

Table 2

Comparison of Categories: Discriminant Function Analysis

Comparison	$D =$	Typical D for (A)	Typical D for (B)	F ratio (df = 3 by 36)
a	$0.90\,X + 0.11\,Y - 0.07\,Z$	9.00	7.36	2.28
b	$0.83\,X - 0.06\,Y + 1.28\,Z$	13.40	6.22	30.86***
c	$1.20\,X - 0.01\,Y + 0.15\,Z$	10.04	2.93	29.38***
d	$1.20\,X - 0.17\,Y - 0.03\,Z$	4.10	0.65	5.39**
e	$1.04\,X + 0.11\,Y - 0.13\,Z$	7.93	3.20	10.64***
f	$-0.15\,X + 0.25\,Y + 0.06\,Z$	4.43	2.72	2.46

** Significant at .01 level.
*** Significant at .001 level.

Table 3

Original Comparisons: Cumulative Percentages Cut Off

Value of *D*	Comparison									
	b		*c*		*d*		*e*		*f*	
	B*	MD†	B	BD‡	D§	MD	D	BD	MD	BD
17 to 18	100									
15 to 16	80									
13 to 14	60	100	100							
11 to 12	45	95	90				100			
9 to 10	30	80	55	100			90			
7 to 8	15	65	25	95	100		55	100		
5 to 6	10	55	15	85	75		25	95	100	100
3 to 4		45		65	50	100	15	60	40	75
1 to 2		30		50	30	90		40	15	60
−1 to 0		5		30	10	25		20		20
−3 to −2		5		10	10	25				
−5 to −4					5	10				

* Bright sample.
† Mental defective sample.
‡ Brain-damaged sample.
§ Dull sample.

For the purpose of cross validation the three tests were repeated on fresh samples (partly drawn from adjacent hospitals) of 20 patients for each category chosen so as to correspond to the original samples in respect to age, sex, and Wechsler vocabulary level. The only divergence in neurological classification was that the case of head injury in the original sample has as its analogue in the second sample a case of tumor in the corpus callosum. As before, the test results were scored by the writer and independently by another person who was provided only with the actual results and scoring instructions; disagreement once again was trivial. The cumulative percentages cut off by different *D*

Table 4

Cross-Validating Comparisons: Cumulative Percentages Cut Off

Value of *D*	Comparison									
	b		*c*		*d*		*e*		*f*	
	B*	MD†	B	BD‡	D§	MD	D	BD	MD	BD
17 to 18	100									
15 to 16	85									
13 to 14	70	100	100							
11 to 12	65	90	90							
9 to 10	55	90	65				100			
7 to 8	35	90	40	100			95	100		
5 to 6	10	80	15	90	100		45	90	100	100
3 to 4		65		70	95		10	65	50	95
1 to 2		35		50	55	100		45	25	50
−1 to 0		25		20	20	75		20	5	25
−3 to −2		15			5	35				
−5 to −4						20				

* Bright sample.
† Mental defective sample.
‡ Brain-damaged sample.
§ Dull sample.

values in the five relevant comparisons of the cross-validating samples are set out in Table 4. It will be noted that the correspondence between the original and the cross-validating samples is close.

Discussion

The fact that mental defectives are significantly discriminated from the bright and dull patients indicates the dependence of successful performance in these tests on general intellectual efficiency. The significant discrimination of the brain-damaged from the bright and dull patients without brain damage suggests that the Strauss-Lehtinen battery, if the Dot-Perseveration Test is dropped and a strict quantitative scoring system is employed for the other three tests, can be of assistance in this problem. The suggestion is strongly confirmed by the cumulative percentages data on the original and cross-validating samples. It is disappointing that there is no significant discrimination between mentally defective and brain-damaged patients, since no test has yet been devised which effectively answers this problem, and the statements of Strauss and Lehtinen (1) about endogenous and exogenous mental defectives have raised hopes that their battery might provide a solution. However, consideration of the cumulative percentages data on this comparison in Tables 3 and 4 indicates that at certain values of D (e.g., 1 to 2) a quite effective discrimination exists, which systematic item analysis might improve.

Summary

The Strauss-Lehtinen Battery for brain-injured children was applied to 80 adult mental hospital patients. The Marble-Board, Figure-Background, and Tactual-Motor tests were found, when quanitatively scored, to give jointly a successful discrimination between brain-damaged patients and bright or dull patients without brain damage. The discrimination between brain-damaged and mentally defective patients was much less satisfactory, but it is suggested that systematic item analysis might improve it. The findings were confirmed on cross-validating samples of 80 closely similar patients.

Received July 16, 1954.

Reference

1. Strauss, A. A., & Lehtinen, L. E. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1947.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Normative Data on Progressive Matrices (1947)[1]

Meredith W. Green and Josephine C. Ewert

Mayo Clinic, Rochester, Minnesota

The Progressive Matrices (3) consist of 36 attractively colored problems in perceptual reasoning that, according to their author, can be presented in either booklet form or as a form board. At the beginning of the series the problems are easy to solve; they gradually become more complex, providing more intricate patterns and requiring more discrimination in making the correct choice.

According to the author, J. C. Raven, the Progressive Matrices, while not a test of general intelligence, will indicate a person's ability to form comparisons and to reason by analogy, as well as to show to what extent he is capable of organizing spatial perceptions into systematically related wholes. The test takes only about 20 minutes, thus affording an efficient means of gaining some information in those clinical situations in which a handicap of either motor coordination or of verbal ability is present.

The norms in Raven's manual are based on the scores obtained by 608 children between the ages of 5 and 11½ years in a rural community of a non-American culture. In our use of this test as part of the psychologic testing of children at the Mayo Clinic, it soon became evident that these norms appeared to be too low.[2] Therefore, we decided to carry out a normative study on children in our area to determine the expected score for various age levels and the relationship existing between the Progressive Matrices and the standard group intelligence tests used in the public schools in our region.

Method of Collecting Data

To facilitate the collection of a large number of test results quickly, the pages of the test booklet were reproduced on individual color slides, which permitted administration to an entire classroom at a time.

This method of administration of the matrices by slides obviously raised the question of the relationship between results of our group tests and results that would have occurred if the test had been given to these children individually. Data were obtained from children between the ages of 6–6 and 9–5; the 30 with the lowest scores on the slide form of the test were retested individually and 45 students were given the individual form prior to the group tests. The results of these variations in testing methods indicate a consistent tendency for the children to show a significant gain when retested, apparently as a result of practice. Since the gains were similar for each group administration (approximately 4 points), it was considered that, for the purpose of amassing normative data, the two methods of testing give comparable results. This agrees with Raven, who stated in his manual that there is apparently no real difference in test scores obtained from the booklet or form-board method of his test. The rank-order correlations were between .62 and .87 for these distributions, which are sharply curtailed by age in both cases and by the narrow range of scores for those who had group tests prior to individually administered tests. This indicates that students tend to maintain their relative position within the group despite the effect of practice.

---

[1] We are indebted to the Right Reverend Monsignor Louis D. O'Day, of St. John's Grade School, the Reverend R. J. Jansen, of St. Francis Grade School, Rochester, Minnesota, Superintendent H. H. Schaefer, of Stewartville, Minnesota, and Superintendent L. J. Gustafson, of Owatonna, Minnesota, for permitting this study to be conducted in their schools.

[2] In a letter to the authors, Raven explained that, while the sample of Dumfries school children was carefully compiled, and there is no reason to doubt its accuracy, he has the impression from repeated evidence collected by others that the intelligence of these Dumfries children is, on the average, lower than that for most of the urban districts of Britain. He also pointed out that there has been for several generations a steady movement of the population away from the rural districts into the cities or abroad.

Our classroom testing included all the children in the second through sixth grades of four different elementary schools in the area of Rochester, Minnesota. The first two were parochial schools located in Rochester; they are estimated to have over 80 per cent of the local Catholic children. The third school was a consolidated one, located in a town of approximately 1,100 population; the fourth school included three-fifths of the children in a town of 10,000.

Results

The slide administration yielded scores for 1,214 children between the ages of 6–0 and 12–5 years. A comparison indicated no significant difference in scores for the children in any particular school. The product-moment correlation between age and score on the Progressive Matrices was .55 (σ_r, .02), indicating that in this group of children a positive and moderately high relationship existed between age and success on this test.

Table 1 shows by six-month intervals the number of children tested, the range of scores on Progressive Matrices, the mean score, the standard deviation, and the median. Also the Dumfries median score is quoted from Raven's manual. It is obvious that the median scores obtained on the 1,214 children from the Rochester area are consistently higher than those given for the Dumfries school children.

While Table 1 indicates that the median scores for children from the Rochester area show a steady increase that parallels increasing age, it is also evident that progression of scores is somewhat erratic. This irregular increase in median scores, together with the small amount of variation in median scores at times, seriously curtails confidence in the interpretation of individual scores at a definite age level. Both of these criticisms of the test, namely the need for more items so that chance errors would not distort the results and the ceiling effect at older ages, have been noted by other authors who have worked with the matrices.

Table 1

Scores for 1,214 Children on Progressive Matrices (1947)

Age (year/month)	Number of children	Mean score	*SD*	Range	Median	Dumfries median*
6/0–6/5	27	18.6	5.5	10–30	17	15
6/6–6/11	48	17.8	5.1	7–27	18	15
7/0–7/5	93	20.2	5.3	10–32	21	16
7/6–7/11	101	21.0	6.3	4–34	21	17
8/0–8/5	110	23.5	5.9	7–34	24	18
8/6–8/11	116	24.4	5.3	8–36	25	20
9/0–9/5	110	25.7	5.0	13–35	26	22
9/6–9/11	118	25.7	5.7	9–35	27	24
10/0–10/5	132	27.7	5.2	9–36	28	24
10/6–10/11	100	28.6	4.5	15–36	29	26
11/0–11/5	107	29.0	4.3	15–35	30	28
11/6–11/11	99	29.7	4.7	10–35	31	
12/0–12/5	53	29.1	5.2	14–35	31	

* Median for children at Dumfries, Scotland.

Our interest then shifted to what relationship existed between scores on Progressive Matrices and the results of standard intelligence tests. As would be expected, the relationship between scores on Progressive Mattrices and Otis mental ages was high; this correlation was .78 for 192 children, all the students in one school between the ages of 6–6 and 12–5. Table 2 presents the correlation found between scores on the matrices and various intelligence tests (1, 2, 4). For various age levels, the correlations between IQ and scores on Progressive Matrices ranged from .19 to .69, but the concentration of correlation coefficients was somewhere near .40.

The frequency of success on each of the 36 items of the Progressive Matrices was computed; data from all children 7–6 to 12–5 years of age in the largest school were used. In Table 3 are recorded performances on the 12 most difficult items, the ones that elicited the smallest number of correct responses. The increase in the proportion of children responding correctly with increasing age was to be expected in view of the moderately high correlation (.55) already noted between age and scores on Progressive Matrices. The fact that these same 12 items were the most

Table 2

Relationships Between Scores on Progressive Matrices (1947) and Results of Otis, California Mental-Maturity, and Kuhlmann-Anderson Tests

Statistic	Age (year/month)					
	6/6–7/5	7/6–8/5	8/6–9/5	9/6–10/5	10/6–11/5	11/6–12/5
Number of children	22	31	51	49	47	36
Correlation with Otis IQ scores	.62**	.69**	.71**	.38**	.19*	.35*
Progressive Matrices, mean and *SD*	19.4± 5.2	20.0± 5.8	25.3± 5.8	27.1± 5.4	29.0± 3.6	30.6± 5.2
Otis IQ, mean and *SD*	113.2± 9.9	110.4± 7.0	114.2±10.5	110.1±10.8	109.3± 9.0	107.0± 8.5
Number of children	32	54	49	50	57	26
Correlation with California Mental-Maturity IQ	.42**	.37**	.42**	.47**	.43**	.45**
Progressive Matrices, mean and *SD*	22.7± 5.8	24.3± 6.5	25.9± 5.3	25.9± 6.0	28.5± 4.4	30.0± 3.6
California Mental-Maturity IQ, mean and *SD*	126.3±12.3	116.6±13.1	115.6±15.0	114.7±14.2	113.8±15.0	108.7±13.4
Number of children		64	72	83	59	61
Correlation with Kuhlmann-Anderson IQ		.28**	.56**	.48**	.52**	.37**
Progressive Matrices, mean and *SD*		22.7± 5.8	24.7± 5.1	28.2± 5.1	29.5± 5.0	29.0± 5.6
Kuhlmann-Anderson IQ, mean and *SD*		107.4± 8.9	101.5± 7.3	102.0±10.1	103.5±10.1	106.0±12.3

* Significant at .05 level.
** Significant at .01 level.

difficult throughout the age range of the children tested and the fact that there was a general tendency for an increase in correct responses to these items with increasing age argue for the consistency of the items.

Table 3

Percentage of Correct Answers on 12 Most Difficult Items of Progressive Matrices (1947) at Various Ages

Item	Age (year/month)				
	7/6–8/5 ($N^*=75$)	8/6–9/5 ($N=75$)	9/6–10/5 ($N=90$)	10/6–11/5 ($N=63$)	11/6–12/5 ($N=63$)
A 11	20	12	27	32	30
12	31	41	61	73	60
Ab 9	45	45	73	78	65
11	48	51	63	72	76
12	8	19	24	33	38
B 6	33	40	65	74	68
7	45	49	68	67	65
8	21	27	49	44	63
9	17	29	53	73	70
10	31	33	63	62	79
11	21	31	50	57	71
12	7	8	20	33	37

* N =number of children included.

Summary and Conclusions

The Progressive Matrices (1947), devised by J. C. Raven, were administered by means of slides to 1,214 children between the ages of 6–0 and 12–5 years in four elementary schools in Rochester, Minnesota, and nearby areas. Preliminary investigation indicated some effect from practice on repetition of the test but no significant difference between scores obtained by the slide form of administration and those obtained by the booklet form of the test.

The median scores for the children of the Rochester area were consistently higher than the median scores reported by Raven in the manual, with a range of differences from 2 to 6 points. Although the difference between the Rochester area median scores at adjacent age levels was frequently merely a point or so, the over-all correlation between scores on matrices and age was moderately high (correlation of .55); a fairly substantial correlation was noted between the Otis mental ages and the scores on Progressive Matrices (correlation of .78). However, correlation with IQ's obtained from three different standard group intelligence tests used in these schools was only moderate (average correlation slightly greater than .40). Scores on Progressive Matrices correlated with the more verbal intelligence tests to about the same degree that they did with those tests which stress nonverbal aspects. Therefore, it would appear that the Progressive Matrices cannot be thought of as a test of nonverbal reasoning ability but instead should be considered as a test of fairly complex intellectual reasoning processes.

The frequency with which the items were answered correctly by a portion of the total group indicated that the items tend to maintain a stable order of difficulty; also, there was a gradual decrease in failure on items over the age range from 6–0 to 12–5 years.

Received July 12, 1954.

References

1. Kuhlmann, F., & Anderson, Rose G. *Kuhlmann-Anderson Intelligence Tests.* Minneapolis: Educational Test Bureau, 1952.
2. Otis, A. S. *Otis Quick Scoring Mental Ability Tests.* Yonkers, N. Y.: World Book Co., 1952.
3. Raven, J. C. *Progressive Matrices (1947).* London: H. K. Lewis, 1947.
4. Sullivan, E. T., Clark, W. W., & Tiegs, E. W. *California Test of Mental Maturity.* Los Angeles: California Test Bureau, 1951.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Relationships Between Verbal and Nonverbal Parts of the CTMM and WISC

Grace T. Altus
Santa Barbara County Schools

One of the reasons for the increasing preference for the WISC (3) over the Stanford-Binet is undoubtedly its separation into Verbal and Performance scales. Similarly, choice of the California Test of Mental Maturity, or CTMM (2), as a group measure of intelligence is frequently made because it yields both Language and Non-Language IQ's. The tacit assumption is often made, as by school psychologists using both tests, that the corresponding verbal and nonverbal portions of the two tests are highly correlated. Some data are here presented bearing on the validity of this assumption.

The subjects were 100 elementary school children, 71 boys and 29 girls, referred to the Guidance Department of the Santa Barbara County Schools by their teachers. The modal reason for referral was consideration for classes for the mentally retarded ($N = 36$), while the remaining cases were referred for a variety of reasons including grade placement, academic disabilities, emotional disturbances, and the like. All children were given the WISC by the writer and the Elementary Short Form of the CTMM in their regular classrooms as part of a routine county testing program. In the cases of 66 children the WISC and CTMM were given less than a year apart, while in no case were the tests as much as three years apart.

Table 1 describes the composition of the group with respect to obtained IQ's on the two tests. It can be seen—as might be expected from the method by which subjects were selected—that the sample is on the average less intelligent but also more variable than an unselected population. The similarity of means on corresponding parts of the two tests is noticeable.

Although the wide sigmas for both tests would tend to raise intercorrelations spuriously, there were other factors operating to reduce such r's: the fact that the group tests were given and scored by teachers in an informal school setting rather than under precise experimental conditions, the time intervening between tests, the inclusion of children with serious reading problems who would be handicapped on the group test, and the like.

Nearly half the children ($N = 49$) had mixed language backgrounds, primarily Spanish-English. Somewhat surprisingly, the majority of this bilingual sample did better on the group test than on the individual one on both verbal and nonverbal sections, with the trend reversed for the unilingual sample. Within both the WISC and the CTMM, however, the bilingual group did conspicuously better in the nonverbal portions than on the verbal portions.

Table 1

Means and Standard Deviations of WISC and CTMM Intelligence Quotients

Test	Mean	*SD*
WISC		
Verbal	82.7	17.2
Performance	89.8	16.3
Full Scale	84.5	17.4
CTMM		
Language	81.8	17.3
Non-Language	88.9	20.0
Total	84.8	17.4

Table 2

Intercorrelations of WISC and CTMM Intelligence Quotients

	WISC		CTMM		
Test	Perf.	Full	Lang.	Non-Lang.	Total
WISC					
Verbal	.77	.93	.71	.65	.76
Performance		.92	.57	.67	.68
Full Scale			.70	.68	.77
CTMM					
Language				.57	.89
Non-Language					.85

Thus, 76% of the bilingual group had higher IQ's on the CTMM Non-Language than Language test, with 27% of the group exceeding an IQ discrepancy of 20 points in this expected direction. These findings would appear to conform to what one might theoretically expect for the test pattern of a bilingual—and in this case also economically underprivileged—group: that is, evidence of greater ability on tests not requiring verbalization. Similar findings for the WISC were reported by the author elsewhere (1).

Table 2 shows the interrelationships derived from Pearsonian correlations of the individual and group tests and their subportions for the entire sample of 100 children. Whereas the corresponding verbal portions of the two tests are substantially correlated, and the corresponding nonverbal portions only slightly less so, it would appear that the CTMM Non-Language has considerably less overlap with its partner scale, the CTMM Language, than the WISC Performance has with the WISC Verbal. The analyses of the two tests, then, cannot be considered fully parallel. In the absence of external validation of the two nonverbal tests, one cannot be sure what the difference represents in behavioral correlates or aptitudes, but surely it is a difference.

Further analysis of test scores was made by tabulating IQ differences for individual children on the three pairs of corresponding IQ's—verbal, nonverbal, and total—of the WISC and CTMM. The nonverbal tests showed the greatest number of wide discrepancies, with one child in four showing a difference in excess of 18 IQ points between WISC Performance and CTMM Non-Language IQ. The difference between *total* IQ's on the two tests was, however, less than 10 IQ points in 65% of the cases, indicating reasonably adequate individual prediction from one test to another.

In short, it would seem justified to conclude that, within a comparable school referral setting, the WISC and CTMM are markedly comparable as to group assessment and roughly comparable as to individual scores and major breakdown into verbal and nonverbal abilities. The precise differentiation of the latter abilities deserves further exploration, could studies be devised to determine the actual meaning of nonverbal as opposed to verbal tests. The fact that bilingual children did better on the respective nonverbal tests than on the verbal appears to be evidence of their intrinsic validity. In any event, the saturation of verbal factors in virtually all the tests herein discussed would imply their usefulness in a school setting.

Received August 26, 1954.

References

1. Altus, Grace. WISC patterns of a selective sample of bilingual school children. *J. genet. Psychol.*, 1953, **83**, 241–248.
2. Clark, W. W., *et al. California Short-Form Test of Mental Maturity; manual.* (Elementary 1950 S-Form.) Los Angeles: California Test Bureau, 1950.
3. Wechsler, D. *Wechsler Intelligence Scale for Children; manual.* New York: Psychological Corp., 1949.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Food Aversions: Some Additional Personality Correlates

Wendell I. Smith, Elizabeth K. Powell
Bucknell University

and Sherman Ross
University of Maryland

One of the interesting aspects of human behavior that has been relatively unexplored is the area of food habits. There are many facets of research which need to be undertaken to assay the variables that enter into this complex universal behavior (8). Foods are a readily available portion of the behavioral environment at which responses can be directed. A series of studies which deal with food aversions has been reported. In some of these (3, 11, 12) neurotic *S*s have been found to select a larger number of foods which they "dislike and would not eat" than have normal controls.

We recently reported on the relationship between scores on the Taylor Manifest Anxiety Scale and food aversions in a college student population (10). A low positive (but significant) correlation was found between scores on the *A* scale and the number of food aversions. Sampson and Bindra (9) have questioned the meaningfulness of high scores on the *A* scale when used with neurotic *S*s. They find that scores in the second and third quartiles were more predictive of neuroticism than scores in the top quartile. These findings may explain the low degree of relationship which we found.

The present study stems from our continued interest in the relationships between food aversions and personality characteristics. We hope to determine the relationship between some additional personality factors and performance on a check list of foods.

If a large number of food aversions is related to neurotic trends, if the overriding characteristic of the neurotic is fear and anxiety, and if the fear and anxiety can be assumed to displace to several areas of behavior, then we can predict that (*a*) *S*s having a large number of food aversions will have marked neurotic characteristics, and (*b*) such *S*s will tend to report that they dislike and will not eat foods which they have never tasted. The second prediction could represent a socially acceptable way of expressing *S*s' fear and anxiety.[1] More simply, those *S*s who have many aversions for foods which they have tasted will generalize this aversion to foods which they have never tasted.

Method

A check list of 28 selected foods (see Table 2) was given to 235 undergraduate students (117 male and 118 female) in introductory psychology classes. The following instructions accompanied the check list:

> Indicate which of the foods in the list below you refuse to eat. Do this as follows: Read over the list slowly. When you come to the name of a food you dislike, put a circle around the number of that food. After marking all the foods you do not like, go back over the list and draw a circle around the *names* of all the foods which you dislike, but have never tasted. After each food you dislike, state the reason for your dislike.

Parts of three paper-and-pencil inventories which purport to evaluate characteristics which we believed might be related to the number of food aversions were also administered to this group. The inventories used were: P-S Experience Blank, Part 1 Physiological (5); The Personal Audit, Form LL, Part V (Stability-Instability) and Part VII (Emotionality-Steadiness) (1); and the F scale of authoritarianism (2). Additional information concerning *S*s' sex and age, father's occupation, religious preference, frequency of attending church, etc., was requested. Names or other identification was not required.

---

[1] We are indebted to Dr. J. W. Gustad, University of Maryland, for assisting us in developing this approach.

Results and Discussion

Two food aversion scores were computed for each *S*. The first (FA-I) was the total number of foods which *S* checked as disliked. The second score (FA-II) was FA-I minus the number of foods checked as disliked but which *S* had never tasted. FA-II then is the number of foods which *S* has tasted and for which he has an aversion. The results obtained from this procedure should be fairly close if *S*s have very few aversions for foods which they have never tasted. The relative magnitude of the difference between FA-I and FA-II should provide information on whether or not familiarity with a food is an important determinant of the number of food aversions (3).

From the total sample of *S*s we drew two samples: High FA and Low FA. The cutoff for the High group was 6 or more checked items, while the cutoff for the Low group was 1 or 0 items. Two such groups were made for each of the FA scores as follows: (*a*) FA-I High ($N = 51$, Mean $= 7.2 \pm 1.4$) and FA-I Low ($N = 51$, Mean $= 0.6 \pm .5$); (*b*) FA-II High ($N = 31$, Mean $= 9.8 \pm 2.3$) and FA-II Low ($N = 40$, Mean $= 0.6 \pm .5$). The mean score for each of the groups on each inventory was computed and tests of the significance of the difference between the High and Low FA groups on each test were made. These results are presented in Table 1.

It is clear from Table 1 that the High and Low groups differed significantly on all but the F scale. In all cases, the mean scores for the High FA groups tended more toward those obtained for neurotic populations than did the mean scores for the Low FA groups.

Interestingly enough, the results on the authoritarian scale for both the High and Low FA groups were nonsignificant even though the description of the authoritarian personality would lead one to expect that the F scale might differentiate between those *S*s who are characterized by relatively high anxiety, relatively poor emotional control, instability, feelings of inadequacy, and physiological dysfunction, and those who are normal in these characteristics. Masling (7) has made this point previously. That the F scale scores of very extreme groups (extreme in that they reported not more than two aversions or nine or more aversions) may differ significantly is indicated by the results which we obtained from a comparison of the 15 highest- and 15 lowest-scoring *S*s. These *S*s were all females. A similar analysis for 30 extreme male *S*s did not result in a significant difference for F-scale scores. This adds to the evidence to be presented that there are marked sex differences in the number of food aversions.

In our earlier study we proposed that the data obtained from a check list of foods might be more meaningful if information on experi-

Table 1

Personality Test Results for High and Low FA Groups

	FA-I				FA-II			
	High		Low		High		Low	
Score	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Emotionality	56.6*	11.9	33.3	11.6	35.9*	11.7	29.9	12.2
Stability	41.2*	22.4	31.0	18.7	45.4*	21.6	30.7	20.1
P-S inventory	119.7*	42.3	90.2	45.0	118.0*	43.6	79.1	51.9
F scale	−7.9	21.5	−5.7	22.4	−9.0	21.0	−4.5	23.9

* Difference between High and Low groups significant at the .01 level of confidence.

ence with each food and the reasons for disliking each food were available. We made an attempt in this study to collect these data by (*a*) having each *S* signify which foods he had never tasted and (*b*) having each *S* state the reason for his aversion to each food checked. From our analysis, it appears that the two food aversion scores have approximately the same discriminative power. Low and High groups whether selected by FA-I scores or by FA-II scores were significantly different on three of the four inventories. This finding was derived from analysis of the data without regard to sex. As we will now show, when the analysis is made taking sex into account, the two food aversion scores for females are significantly different.

Comparisons were made between male ($N = 117$) and female ($N = 118$) *S*s on each of the inventories as well as on food aversions. Only on the trait of stability-instability was the difference between the sexes significant. The female *S*s as a group were somewhat less stable (CR = 3.78). As in previous studies, female *S*s reported more food aversions than did male *S*s. For FA-I, the mean for male *S*s was 3.1 ± 1.7 aversions and for female *S*s it was 5.3 ± 2.9 aversions. For FA-II the mean for male *S*s was 2.6 ± 2.7 aversions and for female *S*s 4.3 ± 2.6 aversions (CR = 2.22). On both scores female *S*s were significantly higher than males. Further, the FA-I mean for females is significantly higher (CR = 2.89) than the FA-II mean, while for males there is no significant difference between the mean FA scores. Seemingly, for our sample of females there was a tendency to dislike foods which had never been tasted, indicating generalization.

Table 2

Percentages of Male and Female *S*s ($N = 235$) Reporting Food Aversions for Each Food Item

Food	Male	Female
Mushrooms	16	26
Cottage cheese	22	19
Swiss cheese	11	17
Tea	6	21**
Bean soup	6	21**
Salmon	9	13
Lima beans	11	14
Radishes	6	11
Cabbage	8	14
Cantaloupe	3	4
Potato soup	9	16
Grapefruit juice	3	9*
Veal chops	2	2
Fried eggs	3	14**
Tomatoes	0	3*
Cherries	0	6**
Pears	0	3*
Salt pork	9	11
Liver	21	37**
Kidneys	18	25
Brains	14	14
Gravy	0	3*
Chili	4	15**
Chocolate	1	1
Beer	12	32**
Buttermilk	45	49
Onions	8	13
Mush	13	19

* Sex difference significant at the .05 level of confidence.
** Sex difference significant at the .01 level of confidence.

A further check for this factor was made by comparing those *S*s who reported aversions to foods they had never tasted with those *S*s who reported that they had tasted each of the foods for which they had an aversion. This was done for female *S*s only. Fifty-six of the females reported one or more aversions to foods never tasted. The mean FA-I score for this group ("generalizers") was 6.5 ± 1.7. A group of 25 *S*s was drawn from the remainder of the female sample on the basis of their FA-I scores. The mean for this group ("nongeneralizers") was 6.1 ± 1.4. Thus, the total number of aversions was held constant and a chi-square test was made to determine whether or not the two groups differed in their means for each of the personality measures. A chi square of 9.575 (3 *df*, $p < .02$) was obtained. The groups differed significantly only on the F scale, on which the nongeneralizers were found to be much less democratic than the generalizers. When the generalizers are compared with the full sample of female *S*s, the mean FA-I score and all inventory scores are practically identical. These results indicate that, although more female *S*s than male *S*s report aversions for foods which have never been tasted, the female *S*s so reporting do not differ in test performance from the total female sample. No ready explanation occurs to us for the lower F-scale score of those *S*s who did not report aversions to foods never tasted.

In Table 2 the sex differences in aversions for each of the 28 foods are presented. Significantly more female *S*s than male *S*s disliked 11 of the foods. When we compare the present findings with those obtained in our previous study, the differences between the sexes overlap on only two foods, chili and beer, indicating that the two samples differed in the frequency with which each food was disliked.

In order to clarify the differences between the extremes of our sample, the 15 *S*s having the highest FA-II scores and the 15 *S*s having the lowest FA-II scores were selected for comparison. Since the High *S*s were all females, only females were selected for the Low group. The Low group mean was 1.3 aversions, while the High group mean was 10.8 aversions. The mean inventory scores for the Low group were well within the range for "normals" while the scores of the High group bordered on those typical for neurotics. All the differences between the two groups were statistically significant as determined by chi-square tests comparing obtained scores on each inventory with expected scores on each inventory. The chi-square values (with 1 *df*) obtained from comparing these Low and High female groups on each test were: P-S Experience Blank, 80.5 ($p < .001$); Stability-Instability, 15.6 ($p < .001$); Emotionality-Steadiness, 14.7 ($p < .001$); and F scale, 4.12 ($p < .05$).

Examination of the biographical data for these 30 *S*s did not reveal any apparent differences between the groups in age, parents' education, religious preference, father's occupation, size of home community, and regularity of attending church. There were two and one-half times as many only children in the High group but the sample is obviously too small to attach great significance to this factor.

When *S*s' reasons for their aversions are considered, one is struck with the superficiality of the answers. After examining the responses, we decided against any formal categorization. Slightly less than half of all answers consisted of "I don't like the taste." This response leaves us little more information than we had when we started. A comparison of the reasons stated by the Low group and by the High group on FA-II scores provides only one interesting point. Forty per cent of the reasons given by the High group could be categorized as responses which would probably not usually characterize the food. For example, such adjectives as "mushy" and "slimy" were applied to foods such as liver, tomatoes, beans, and chili. Such descriptions as "it gives me indigestion," "I don't like the feel of it," and "I can't even stand the thought of it" were common. These statements were applied to such commonly consumed items as beans, tomatoes, liver, cabbage, cherries, eggs, onions, and beer. This point is of interest, since only one response in the Low group could be so categorized. This single response was given to fried eggs. The *S* wrote: "I am allergic to concentrates of protein." It is unfortunate that our *S*s were anonymous, since it might have been valuable to interview them intensively concerning the reasons for their aversions.

The results of this study give indirect support to the hypothesis that there is some generalization of affect to the socially accepted area of food habits. There is a tendency for individuals who report disturbances in affective behavior to report an unusual number of food aversions.

Lewin's (4) contention that "people like what they eat, rather than eat what they like" seems to be an insightful inference. The *S*s in our study did not like those foods with which they had had no experience. Familiarity of the food appears to be an important factor.

The bases for the aversions reported still remain to be clarified. Intensive interviewing might result in fresh leads to this problem. An interesting lead to be explored in future studies is supplied in Margaret Mead's statement that food habits may be influenced by: "The Puritan tradition of a connection between food which is disliked, and the tendency in communities with a Puritan tradition to use food for purposes of reward and punishment . . ." (7, p. 23).

Summary

This study deals with the relationships between food aversions (FA) and personality characteristics. A group of 235 undergraduate students was studied for possible relationships between FA and scores made on four selected personality inventories. These inventories were: (*a*) P-S Experience Blank, (*b*) The Personal Audit (Stability-Instability), (*c*) The Personal Audit (Emotionality-Steadiness), and (*d*) F scale of authoritarianism.

Two kinds of FA scores were computed and used to categorize the sample in a variety of ways for comparison purposes. The first score was based on the total number of items checked as disliked from the list of 28 foods given. The second score was based on the number of items checked as disliked but which *S* reported he had tasted. The basic comparisons were made between selected High and Low groups for each of these scores, and for sex differences using the total sample.

The results permit the following conclusions:

1. High and Low FA groups on each scoring technique differed significantly on all of the personality inventories, except for the F scale. The High groups were closer to the scores obtained by neurotic *S*s on these inventories.

2. Female *S*s ($N = 118$) reported significantly higher FA scores than male *S*s ($N = 117$).

3. No significant differences were found between the two FA scores of the male *S*s. A significant difference was found between the two FA scores for female *S*s. This result was interpreted to mean that there was a tendency for female *S*s to dislike foods which they had never tasted.

4. The stated reasons for an FA by *S*s were most frequently in the form of descriptions which would probably not be the usual characterization of the food.

Some limited interpretations were made of these findings in regard to the role of foods as socially acceptable objects against which the effects of anxiety and fear can be directed.

Received July 6, 1954.

References

1. Adams, C. R., & Lepley, W. M. *The Personal Audit* (Forms L and S). Chicago: Science Research Associates, 1941.
2. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.
3. Gough, H. G. An additional study of food aversions. *J. abnorm. soc. Psychol.,* 1946, **41**, 86–88.
4. Lewin, K. Forces behind food habits and methods of change. In C. E. Guthe [Chm.], The problem of changing food habits; report of the Committee on Food Habits. *Nat. Res. Coun. Bull.,* 1943, No. 108. Pp. 35–65.
5. McFarland, R. A., & Seitz, D. P. A psychosomatic inventory. *J. appl. Psychol.,* 1938, **22**, 327–339.
6. Masling, J. M. How neurotic is the authoritarian? *J. abnorm. soc. Psychol.,* 1954, **49**, 316–318.
7. Mead, Margaret. In C. E. Guthe [Chm.], The problem of changing food habits; report of the Committee on Food Habits. *Nat. Res. Coun. Bull.,* 1943, No. 108. Pp. 20–31.
8. National Research Council. Committee on Food Habits. Manual for the study of food habits. *Nat. Res. Coun. Bull.,* 1945, No. 111.
9. Sampson, H., & Bindra, D. "Manifest" anxiety, neurotic anxiety, and the rate of conditioning. *J. abnorm. soc. Psychol.,* 1954, **49**, 256–259.
10. Smith, W. I., Powell, Elizabeth, & Ross, S. Manifest anxiety and food aversions. *J. abnorm. soc. Psychol.,* 1955, **50**, 101–104.
11. Wallen, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.,* 1945, **40**, 77–81.
12. Wallen, R. Food aversions in behavior disorders. *J. consult. Psychol.,* 1948, **12**, 310–312.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Effect of Pinhole Observation on the Flicker Fusion Frequency in Patients with Brain Pathology[1]

James C. Stauffacher, John B. Marks
VA Hospital, American Lake, Washington

and Albert F. Ax
University of Washington

This study was designed to determine whether or not the lowered critical flicker fusion frequency (CFF) of patients with brain pathology is a result of alterations in the size of the pupil and whether use of a restricted pinhole pupil might not increase the effectiveness of CFF examinations in screening patients with brain pathology.

The experimental group consisted of 21 patients with various organic brain disorders. The control group consisted of 21 psychotic patients with no known brain pathology. Each patient of the brain pathology group was matched in age with a control patient.

The intermittent visual stimuli were obtained by using a Strobotac to activate a neon bulb which was painted black except for an orifice ⅜ inch in diameter, through which the glow plate was viewed. The bulb was mounted at one end of a lightproof box 40 inches from the viewing cowl from a stereoscope. One eye could be shielded by the insertion of a plate, and a second plate with a pinhole of .046-inch diameter could also be inserted before the other eye.

Each patient was tested for CFF under three conditions: with the usual binocular stimulation, monocular stimulation, and monocular stimulation through the pinhole. The order of these conditions was alternated systematically and the establishment of a "time set" was avoided by systematic alteration of the beginning frequency of the light stimulus. Fatigue was minimized by having subjects close their eyes between observations.

The usual binocular condition discriminated between the brain pathology and control groups at the .01 level of confidence when the mean of the up trials (28.05 and 31.18), the mean of the down trials (27.04 and 31.30), and the mean of the up and down trials (27.55 and 31.25) were considered. When compared with binocular observation, the monocular condition produced slightly smaller differences between the pathology and control groups and the difference between the groups reaches the .01 level of confidence for each criterion used. When pinhole observation is used, the pathology group has lower CFF in mean up trials (17.54 and 18.20), mean down trials (15.94 and 17.89), and mean of up and down trials (16.75 and 18.04). However, the differences are smaller and the reliability is greatly reduced, the significance levels being .25, .01, and .05 respectively.

These results indicate that differences in CFF between pathology and normal groups cannot be explained in terms of variations in pupil size. The usual binocular observation yielded a more reliable difference between pathology and control groups than did the pinhole observation used. The factors responsible for the reduction in the reliability of pinhole observation are not indicated from the study.

Brief Report
Received November 23, 1954.

---

[1] An extended report of this study may be obtained without charge from J. C. Stauffacher, VA Hospital, American Lake, Wash., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4467 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Emotional Adjustment and Dieting Performance[1]

John Summerskill and C. Douglas Darling
Department of Clinical and Preventive Medicine, Cornell University

Two recurring observations on obesity suggest its relationship to psychosomatic disorders. First, body weight is directly related to caloric intake (3, 4, 5, 9). Secondly, caloric intake is in some degree and in some manner related to the individual's emotional state (1, 2, 3, 6, 7, 8, 10). Consequently, the treatment of obesity necessarily involves dieting, but treatment is successful only if emotional conditions are favorable. This research was focused on the latter relationship: emotional influences on dieting performance.

The immediate problem was posed in experiments conducted at Cornell University on the efficacy of certain diets (11, 12). Subjects were given three carefully measured meals daily for three months. Evaluation of dietary effects involved continuing weight checks, basal metabolism studies, fecal balance studies, blood sample analyses, etc. The validity of the findings from these intensive investigations was largely dependent upon the ability of the subjects to adhere rigorously to the diet regime for a full three-month period. From this standpoint it was important to select subjects who were able to diet successfully.

The authors hypothesized that there is a positive relationship, subject to quantification, between emotional stability and dieting success. Validation of a technique for selecting dietary research subjects was undertaken on the assumption that it would be relevant to the diagnosis and treatment of obese patients in clinical settings. Suitability for dieting is important in view of the recognized possibility that acute disturbance can be precipitated in maladjusted individuals when eating habits are radically altered (2).

Method

Psychological tests were administered to subjects of three dietary experiments conducted in the spring of 1952, 1953, and 1954. The total number of subjects tested was 29, including 17 college women and 12 college men with an average weight excess of 25%. Scores on emotional adjustment were subsequently compared with: (*a*) pounds lost through dieting and (*b*) diet performance ratings. The second criterion was included in the design because pounds lost does not adequately represent dieting success as it relates to individual differences in percentage of overweight, activity expenditure, and other factors. Each dieter was rated on his demonstrated acceptability (+ or −) for controlled dieting by the physician, nutritionist, or dietician who had supervised the dieting. Two independent ratings were available for 23 subjects and these were in agreement in all cases but one. Specific methods employed in each of the three experiments were as follows:

Experiment I (*1952*). During the course of the first dietary experiment the nutritionist learned that not all subjects were adhering strictly to the diet provided. At the completion of dieting, this situation was brought to the attention of the writers who then studied the six of ten subjects still on campus. Each subject was seen in an evaluative psychiatric interview and was administered a battery of psychological tests including: (*a*) Bell Adjustment Inventory, (*b*) Bernreuter Personality Inventory, (*c*) MMPI, and (*d*) Thurstone Personality Schedule.

---

[1] The diet experiments, conducted by the School of Nutrition, Cornell University, were supported in part by the Cornell Nutrition Fund and in part by the Gannett Nutrition Fund and in part by an appropriation from the State of New York through the State University. The authors wish to express their appreciation to Norman S. Moore, M.D., Jeffrey Fryer, M.D., and Charlotte M. Young, Ph.D., of the School of Nutrition, for their interest and cooperation in the psychological phase of the research.

Experiment II (1953). The results of Experiment I suggested that prediction of dieting performance was feasible from a single, objective index of emotional adjustment. There were also indications that adjustment during dieting was largely expressed through psychosomatic complaints. Accordingly, the 11 subjects of the second dietary experiment were tested before dieting began with the: (*a*) Bell Adjustment Inventory and (*b*) Psychosomatic Inventory. These tests were repeated at the midway point and at conclusion of the diet.

Experiment III (1954). To cross-validate the selection procedure and increase N the two tests employed in Experiment II were administered to 26 applicants for the third dietary experiment. Twelve subjects were then selected for the diet, each above average on test measures of emotional stability.

Results

A positive relationship between above average emotional adjustment and dieting success was found within each experimental group and for the experimental groups combined. Table 1 shows the dieting performance of subjects who demonstrated better than average vs. average or below average emotional adjustment according to the college norms of the Bell Adjustment Inventory. Considering all experimental groups, 78% of the well-adjusted individuals were independently rated as acceptable dieters whereas only 18% of the other individuals were so rated. This relationship is highly significant ($p >$ .01) according to a chi-square test on the fourfold table (Yates' correction for small frequencies employed). With regard to the other criterion, well-adjusted individuals lost an average of 21.0 lb. ($SD = 5.25$) whereas other individuals lost an average of 13.6 lb. ($SD = 5.17$). A t test shows that this difference is highly significant ($p >$ .01).

Table 1 contains additional data on the predictive value of the emotional scale of the Bell Adjustment Inventory. Dieters selected for Experiments I and II without the use of test scores proved to be acceptable in 7 of 17 cases. On the other hand, dieters selected for Experiment III on the basis of test scores proved acceptable in 9 of 12 cases. In the three groups test scores accurately classified dieters as acceptable or unacceptable in 23 of 29 cases.

Other observations, some based on numbers too small for statistical treatment, were as follows:

1. Psychological adjustment scores from the P-S Inventory (23 cases) were in essential agreement with Bell emotional scores and showed a similar positive relationship with dieting performance.
2. Psychiatric interview findings on emotional stability (6 cases) were in essential agreement with test indices of emotional adjustment.
3. Indices of adjustment in other areas (e.g., the Bell home, health, and social adjustment scales—29 cases) were not related to dieting performance.
4. Test scores before, during, and after dieting (11 subjects) were essentially unchanged, with one exception. The number of somatic symptoms reported was somewhat greater during dieting.

Table 1

Diet Performance Ratings and Mean Weight Loss of Dieters with Above Average and Average or Below Average Bell Emotional Adjustment Scores

	Dieters with above average Bell emotional adjustment scores			Dieters with average or below average Bell emotional adjustment scores		
Experimental group	N	Number rated acceptable	Mean pounds lost	N	Number rated acceptable	Mean pounds lost
Experiment I (1952)	2	2	22.5	4	1	14.4
Experiment II (1953)	4	3	17.5	7	1	13.1
Experiment III (1954)	12	9	21.9	0		
All groups	18	14	21.0	11	2	13.6

5. Analysis of Bell emotional adjustment items showed no consistent pattern of symptoms among maladjusted dieters (11 cases). Similarly, there was no single or unique personality pattern evident in the psychiatric interviews of three dieters of inferior emotional stability.

Discussion

It is recognized that results are based on a small sample of college students and that extension of conclusions to a general population will require further research. Confidence in the demonstrated relationship between tested emotional adjustment and dieting performance in college students is strengthened by the consistency of findings in three discrete experiments. With this in mind, the writers feel that objective tests of emotional adjustment, e.g., the Bell or P-S Inventory, can be profitably used in selecting individuals who can handle the pressures involved in controlled dieting.

The findings also support the authors' clinical observation that serious emotional disturbance may or may not be present with obesity. Psychiatric case studies of obese patients sometimes leave the impression that obesity invariably reflects serious deep-seated conflict. The present results indicate that obesity can be associated with varying states of emotional adjustment from the essentially normal to the seriously disturbed. When signs of good emotional adjustment are present, the chances that dieting will be successful appear to be about four times as great as they are when these signs are absent.

Summary

The purpose of the research was to examine the relationship between emotional adjustment and dieting performance in order to devise a procedure for selecting individuals suitable for controlled dieting. A statistically significant relationship was found between objective test measures of emotional adjustment and (*a*) pounds lost through dieting and (*b*) independent diet performance ratings. The results show that psychological instruments such as the Bell Adjustment Inventory, or Psychosomatic Inventory, can be profitably used in eliminating individuals whose participation in rigorous dieting can have undesirable consequences. The findings also indicate that obesity can be associated with varying states of emotional adjustment from the essentially normal to the seriously disturbed.

Received June 24, 1954.

References

1. Bruch, Hilde. Psychological aspects of obesity. *Bull. N. Y. Acad. Med.*, 1948, **24**, 73–86.
2. Bruch, Hilde. Psychological aspects of reducing. *Psychosom. Med.*, 1952, **14**, 337–346.
3. Hamburger, W. W. Emotional aspects of obesity. *Med. Clin. N. Amer.*, 1951, **35**, 483–499.
4. Newburgh, L. H. The cause of obesity. *J. Amer. med. Ass.*, 1931, **97**, 1659.
5. Newburgh, L. H. Obesity. *Arch. int. Med.*, 1942, **70**, 1033.
6. Reeve, G. H. Psychological factors in obesity. *Amer. J. Orthopsychiat.*, 1942, **12**, 674–679.
7. Rennie, T. A. C. Obesity as a manifestation of a personality disturbance. *Dis. nerv. Syst.*, 1940, **1**, 238–247.
8. Richardson, H. B. Obesity and neurosis. *Psychiat. Quart.*, 1946, **20**, 400.
9. Rynearson, E. H., & Gastineau, C. F. *Obesity.* Springfield, Ill.: Charles C Thomas, 1949.
10. Schick, A. Psychosomatic aspects of obesity. *Psychoanal. Rev.*, 1947, **34**, 173–183.
11. Young, Charlotte M. Weight reduction using a moderate-fat diet. *J. Amer. diet. Ass.*, 1952, **28**, 410–416.
12. Young, Charlotte M., Ringler, I., & Greer, Betty J. Reducing and post-reducing maintenance on the moderate-fat diet: metabolic studies. *J. Amer. diet. Ass.*, 1953, **29**, 890–896.

NEW BOOKS AND TESTS

Books

Bauman, Mary K. (Ed.) *Adjustment to blindness.* Harrisburg, Pa.: Commonwealth of Pennsylvania, State Council for the Blind, 1954. Pp. xiii + 198.

This intensive study of 443 blind persons from six states is a real contribution to the literature of adjustment to handicaps. The three groups studied were: (*A*) employed and generally well-adjusted ($N = 162$), (*B*) well-adjusted except in employment ($N = 150$), and (*C*) generally poorly adjusted, including employment ($N = 131$). They were first found not to differ in sex, color, residence, birthplace, height, weight, or age. (Group *B* was made up of somewhat younger people.) With these factors constant, differences appeared in favor of well-adjusted employed blind in intelligence and personality inventory scores. These factors overshadowed amount of visual loss, health, education, and family and social interaction as influential causes. It would have been more satisfying to the critic to have a clearer discussion of the use of the criteria in setting up and separating the three groups. Although most of the criteria were objectified and tabulated, their importance was minimized by reporting them along with the dependent variable. Since a personality inventory adapted to the blind was administered rather than a standard one, it would have been well to have included it, in addition to a description of its parts and reliability coefficients. This report presents a prodigious amount of reliable information on employment, recreation, home life, travel, and health which will greatly assist those interested in the rehabilitation of the blind. A helpful 123-title bibliography concludes the volume.—*F. McK.*

Holsopple, James Quinter, & Miale, Florence R. *Sentence completion.* Springfield, Ill.: Charles C Thomas, 1954. Pp. xiii + 177. $5.50.

This little book has many charms. First, it is packed with illustrative protocols of sentence completions by persons described in sufficiently relevant detail for the reader either to check his own interpretations, or to determine for himself whether or not the responses really have the flavor of genuine clinical utility. Second, it contains an extensive case interpretation that is a model of teaching by example. Third, the book is thoroughly free from both clinical jargon and extravagant claims. Holsopple and Miale overtly deny that the sentence completion method is a "test," constructed on the model of a yardstick or a clock, and argue that the task involved in using sentence completions is that of sensitive interpretation, not of scoring. This appropriate frankness leads them to emphasize the point that the clinician is actually part of the instrument in the sentence completion technique, as in virtually all projective devices. It leads them away from any attempt to quantify, beyond a voluminous listing of the most frequent completions to their particular stems. As a result, they implicitly raise the two central issues of what attributes characterize a competent clinician and to what extent clinical psychology may accept Thorndike's dictum about the measurability of all things. That Holsopple and Miale do not deal directly with these problems is hardly to their discredit. They have faced them in a context that is instructive, reflective of an intimate and sympathetic knowledge of the clinician's job, and fruitfully suggestive for all clinical psychologists who take their obligations seriously.—*E. J. S.*

Lewis, Hilda. *Deprived children.* New York: Oxford Univer. Press, 1954. Pp. xvii + 163. $1.55.

This is a report of 500 children who entered the Mersham Reception Centre in Kent, England, between October, 1947 and July, 1950. The Mersham Reception Centre was organized to serve two purposes: first, to plan what would be best for the children entering, and second, to gather data that might throw light on factors that were causing the social maladjustment of these children. The book tells not only the condition of the children when they entered the Centre, with their backgrounds, but also gives the results of a follow-up study made two years later. The follow-up study is especially interesting, both as to the methods used in obtaining data and the results.—*B. M. L.*

Lowenfeld, Margaret. *The Lowenfeld Mosaic Test.* London: Newman Neame; and New York: Psychological Corp., 1954. Pp. 360 + 144 color plates (boxed). $10.00.

Lowenfeld's manual for her well-known Mosaic Test deals comprehensively with its nature and administration, the definition and classification of the designs produced, and the clinical implications drawn from its results. The standard directions for adminis-

---

Note.—The reviews were prepared by the Editor and Advisory Editors, who may be identified by their initials.

tration, and especially the definitions and illustrations of the terms used to describe the designs, are useful to both clinical and research workers. At last they now will have a uniform method for communicating about the Mosaic. The 144 color plates make concepts clear that are hard to describe in words or even in uncolored illustrations. In its clinical sections, the book displays honest confusion. Numerous case studies give, by implication more than direct assertion, the "typical" designs produced by patients with many syndromes, children of various developmental stages, and people of several cultures. In discussing generalities, the author sometimes makes glowing claims—". . . the L. M. T. often makes it possible to make a rapid and certain diagnosis in cases of doubt . . ." (p. 246)—and sometimes shows awareness of the limitations of her method—". . . differential diagnosis between a severe degree of neurosis, a low level of inherent intelligence, and severe schizophrenia, is one of extreme difficulty, since in all three cases subjects are apt to . . ." (p. 146). A striking omission, all too familiar in books on new clinical techniques, is even the simplest evidence on validity, other than anecdotes. Is it too much to ask that authors at least count and report the percentages of hits and misses made by the test when predicting certain verifiable diagnoses or other events?—*L. F. S.*

McNemar, Quinn. *Psychological statistics.* (2nd Ed.) New York: Wiley, 1955. Pp. vii + 408. $6.00.

Like the first edition (*J. consult. Psychol.*, 1949, 13, 308), the revision of McNemar's *Statistics* remains a clear and precise resource for instruction and reference. Changes include some expansions and rearrangements to make the text less difficult, a new treatment of interaction in complex analysis of variance, and a short but useful chapter on distribution-free methods.—*L. F. S.*

Meehl, Paul E. *Clinical versus statistical prediction.* Minneapolis: Univer. of Minnesota Press, 1954. Pp. x + 149. $3.00.

Meehl's essay on the clinical and the statistical, of which many of us have heard rumors for several years, proves to have been worth waiting for. It is a superb piece of thinking, at once fair-minded and stimulating, profound and readable, serious and witty. First, he clears away some cobwebs of semantics that have kept clinicians and the statistically minded from thinking together constructively. Second, he analyzes with penetrating sympathy the logic of clinical activity, in a chapter which is perhaps the book's most distinctive contribution. Third, he reviews twenty empirical studies—all that can be found today—which attempt to compare the values of clinical and actuarial prediction. It is something of a shock to find how meager is the evidence. Some of the twenty studies are manifestly imperfect in design, others deal with issues such as academic prediction that most clinicians would regard as peripheral. The studies find in favor of the actuarial method, at least by a shade, but they point clearly to the need for more and better research. Proponents of both the statistical and clinical approaches need to read this little volume. And let neither side misperceive the author as being in the camp of the enemy; he has much to teach all of us. Read and see for yourself.—*L. F. S.*

Rubinstein, Eli A., & Lorr, Maurice. (Eds.) *Survey of clinical practice in psychology*. New York: International Universities Press, 1954. Pp. xvii + 363. $6.00.

The purpose of this volume is to offer "a comprehensive description of the function and role of the clinical psychologist." It does this by getting some 32 contributors from various clinical fields "to discuss the relationship of clinical psychology to the particular setting in which he practices and to describe the functions of the clinician which characterize his operation in that setting or organization." All the typical clinical settings are represented, and the list of contributors reads like a Who's Who of clinical psychology. If we start in alphabetical order with the list of contributors, we find Bobbitt writing on the Public Health Service, Brewer on the Wichita Guidance Center, and Brotemarkle on the University of Pennsylvania Psychological Clinic; and so on! Such compendia are often thrown together hurriedly of second-rate contributions. This one is an exception. It is obvious that both the authors and the editors have put their hearts into their work. The result is excellent. It is of course a fine reference book, but it also makes stimulating reading. This reviewer would label it a "must" for every clinician's bookshelf.—*W. A. H.*

Schafer, Roy. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954. Pp. xiv + 446. $8.75.

Believing that ". . . much of today's most important theoretical and clinical progress in the study of personality is being made through psychoanalytic ego psychology," and that, in the Rorschach, "Score analysis is *minimal* analysis," Schafer proceeds to demonstrate that Rorschach interpretation can be tremendously enriched by the application of the concepts of psychoanalytic psychology. Interwoven throughout the book is a simplified exposition of some of the basic psychoanalytic concepts, but the application of these concepts to Rorschach interpretation is very much deeper. There is an excellent chapter on interpersonal dynamics in the test situation—a subject too often neglected in Rorschach literature—and another worth-while chapter on the theory of the Rorschach response process, also within the framework of psychoanalytic psychology. Rorschach workers will find the discussion on content analysis, called "thematic analysis" by Schafer, rewarding. A large portion of the work is devoted to psychoanalytic defense mechanisms, studied in relation to Rorschach responses; "obsessive-compulsive defense operations" are included among them. The detailed case presentations add substance to these discussions.—*M. K.*

Test

Edwards, Allen L. *Edwards Personal Preference Schedule*. College-adult. 1 form. Untimed, (40) min. Booklet ($3.00 per 25) with template, manual, pp. 18; answer sheet and profile ($2.90 per 50); specimen set (60¢). New York: Psychological Corp., 1953, 1954.

Edwards has presented an ingenious and novel instrument for personality assessment which will surely evoke wide comment, considerable clinical use, and much research. The *Schedule* is designed in terms of 15 of Murray's manifest needs—achievement, deference, order, exhibition, autonomy, etc.—each of which is paired twice with each of the others. The 225 forced-choice items each consist of a pair of alternatives carefully equated for social desirability. As a result of this unusual and valuable matching, only two of the need scores have correlations with social desirability significantly above zero and these two are low (.32). The *PPS* thereby sidesteps that pitfall of many questionnaires, ego involvement. Because the need scales are short, the modest reliabilities are not unexpected: internal consistencies range from .60 to .87, and retest correlations from .74 to .88. The subscore intercorrelations are low. The validity of such a schedule is not easily expressed in simple terms, but the manual contains interesting data on clinical observations, and on relationships with ratings and other questionnaires. It is a long time since this reviewer has seen a questionnaire that seems to possess such potentialities for use and research.—*L. F. S.*

Books Received

Adams, Donald K. *The anatomy of personality*. Doubleday Papers in Psychology, No. 11. Garden City, N. Y.: Doubleday, 1954. Pp. vii + 44 (paper). 85¢.

Allen, Robert M. *Elements of Rorschach interpretation*. New York: International Universities Press, 1954. Pp. 242. $4.00.

Balint, Alice. *The early years of life*. New York: Basic Books, 1954. Pp. viii + 149. $3.00.

Bernstein, Arnold. *On the nature of psychotherapy*. Doubleday Papers in Psychology, No. 12. Garden City, N. Y.: Doubleday, 1954. Pp. ix + 36 (paper). 85¢.

Blum, Lucille Hollander, Davidson, Helen H., & Fieldsteel, Nina D. *A Rorschach workbook*. New York: International Universities Press, 1954. Pp. iv + 167 (paper). $2.00.

Coleman, Lester L. *Freedom from fear*. New York: Hawthorn, 1954. Pp. 285. $3.95.

Cowell, Roberta. *Roberta Cowell's story*. New York: British Book Centre, 1954. Pp. ix + 208. $3.00.

Deutsch, Felix, & Murphy, William F. *The clinical interview*. Vol. 1. *Diagnosis*. New York: International Universities Press, 1954. Pp. 613. $10.00.

Duvall, Evelyn Millis. *In-laws: pro & con*. New York: Association Press, 1954. Pp. viii + 400. $3.95.

Ferm, Vergilius. *A dictionary of pastoral psychology*. New York: Philosophical Library, 1954. Pp. xi + 336. $6.00.

Gondor, Emery I. *Art and play therapy*. Doubleday Papers in Psychology, No. 10. Garden City, N. Y.: Doubleday, 1954. Pp. x + 61 (paper). 85¢.

Hadamard, Jacques. *The psychology of invention in the mathematical field*. New York: Dover Publications, 1954. Pp. xiii + 145. $1.25 paper; $2.50 cloth.

Hall, Calvin S. *A primer of Freudian psychology*. Cleveland and New York: World Publishing Co., 1954. Pp. xii + 137. $2.50.

Heckler, Edwin L. *The meat packing industry*. Vocational and Professional Monographs, No. 46. Cambridge, Mass.: Bellman Publishing Co., 1954. Pp. 23 (paper). $1.00.

Husquinet, Albert. *L'adaptation scolaire et familiale des jeunes garçons de 12 à 14 ans*. Paris: Société de l'Edition "Les Belles Lettres," 1954. Pp. 199 + 17.

Keller, Fred S. *Learning: reinforcement theory*. Doubleday Papers in Psychology, No. 13. Garden City, N. Y.: Doubleday, 1954. Pp. ix + 37 (paper). 85¢.

Kerman, Edward F. *What is electroshock therapy?* New York: Exposition Press, 1954. Pp. 152. $3.50.

Macfarlane, Jean W., Allen, Lucile, & Honzik, Marjorie P. *A developmental study of the behavior problems of normal children between twenty-one months and fourteen years*. Berkeley and Los Angeles: Univer. of California Press, 1954. Pp. vii + 222 (paper). $2.25.

Nottingham, Elizabeth K. *Religion and society*. Doubleday Short Studies in Psychology, No. 5. Garden City, N. Y.: Doubleday, 1954. Pp. x + 84. 95¢.

Robinson, Mary Frances, & Freeman, Walter. *Psychosurgery and the self*. New York: Grune & Stratton, 1954. Pp. ix + 118. $3.00.

Sasha, Adrian Waldo. *The laws of life*. San Francisco: Living Knowledge Foundation, 1954. Pp. 224. $5.00.

Scott, Edward M. *Harvest time*. New York: Pageant Press, 1954. Pp. 137. $3.00.

Stalnaker, John M., & Dykman, Ross A. *Admission requirements, 1955*. Chicago: Association of American Medical Colleges, 1954. Pp. iii + 195. $2.00.

Ulett, George. *Rorschach introductory manual*. Beverly Hill, Calif.: Western Psychological Services, 1955. Pp. 49 (paper). $3.50.

Wolfenstein, Martha. *Children's humor*. Glencoe, Ill.: Free Press, 1954. Pp. 224. $3.75.

JOURNAL OF CONSULTING PSYCHOLOGY

June, 1955 Vol. 19, No. 3

Contents

A Lobotomy Prognosis Scale: *Wesley C. Becker and Robert L. McFarland* - - - - - - 157

Psychosomatic Patients and Other Physically Ill Persons: A Comparative Study: *Sheldon E. Waxenberg* - - - - - - - - - - - - - - - - - - - 163

Parental Figures in Sentence Completion Test, in TAT, and in Therapeutic Interviews: *Mortimer M. Meyer and Ruth S. Tolman* - - - - - - - - - - - - - 170

Some Standardized Scales for Disorganization in Schizophrenic Thinking: *William A. Hunt and Franklyn N. Arnhoff* - - - - - - - - - - - - - 171

Differential Prediction of a Specific Behavior from Three Projective Techniques: *Claire M. Vernier, J. Frank Whiting, and Malcolm L. Meltzer* - - - - - - - - - 175

The Psychiatric Validity of an Inverted Factor Analysis of Rorschach Scoring Categories: *A. W. Bendig and Roy M. Hamlin* - - - - - - - - - - - - - 183

Various Rorschach Indices as Discriminators of Marked and Little Conceptual Disorganization Among Schizophrenics: *Bernard A. Stotsky and James F. Lawrence* - - - - 189

The Separation Capacity of the Rorschach: *Raymond J. Corsini, Winfield E. Severson, Thomas E. Tunney, and Harold F. Uehling* - - - - - - - - - - - - 194

The Effect of the Psychotherapist's Personal Analysis Upon His Techniques: *Hans H. Strupp* - - - - - - - - - - - - - - - - - - 197

Changes in the Self Concept Without Psychotherapy: *Donald M. Taylor* - - - - - - 205

Self-Evaluations, Group Evaluations, and Objective Measures: *Wilse B. Webb* - - - - 210

MMPI Validity Scores as a Function of Increasing Levels of Anxiety: *Joseph D. Matarazzo* 213

The Relationship of Manifest Anxiety to Wechsler-Bellevue Subtest Performance: *Ruth G. Matarazzo* - - - - - - - - - - - - - - - - - - 218

The Relation of Manifest Anxiety to Association Productivity and Intellectual Attainment: *Anthony Davids and Charles W. Eriksen* - - - - - - - - - - - 219

A Failure to Replicate the Finding of a Negative Correlation Between Manifest Anxiety and ACE Scores: *R. E. Schulz and Allen D. Calvin* - - - - - - - - - - 223

Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: Addendum: *Betsy Worth Estes* - - - - - - - - - - - - - - - - - 225

Zaslow's Test of Concept Formation: Reliability and Validity: *Sol Kugelmass and Marvin R. Foundeur* - - - - - - - - - - - - - - - - - 227

The Effects of Serial Position and Design Difficulty on Recall of the Bender-Gestalt Test Designs: *Leonard D. Goodstein, Charles D. Spielberger, John E. Williams, and W. Grant Dahlstrom* - - - - - - - - - - - - - - - - - 230

New Books and Tests - - - - - - - - - - - - - - - - - 235

Bureau of Edn. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated ....
Accs No ....

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

A Lobotomy Prognosis Scale[1,2,3]

Wesley C. Becker and Robert L. McFarland
VA Hospital, Palo Alto, California

Since the development of the prefrontal lobotomy technique by Moniz in 1936 and its introduction into this country by Freeman and Watts, much effort has been devoted to evaluating this therapeutic method. From a psychiatric viewpoint, evaluation has centered on selection of suitable candidates for lobotomy and on assessment of the changes produced by the operation. In this paper we are concerned primarily with the problem of selection.

A review of the literature readily reveals many opinions and some experimental evidence on factors related to improvement after lobotomy. For example, such factors as the tortured self-concern syndrome, rapid onset, lack of deterioration, and good premorbid adjustment have been repeatedly reported as favorable prognostic indicators. The primary methods used in gathering such data are the psychiatric interview and history, illustrated by the studies of Partridge (5) and Greenblatt *et al.* (2); and psychological testing, such as by Scherer's (6) study of memory, abstraction, and motor functioning. Though the literature abounds with signs and some significant differences to guide in the selection of lobotomy candidates, there has been no systematic attempt to quantify the data in the form of a prognosis scale and to test the predictive validity of such a scale. This lack prompted the present study.

This study attempted to construct a prognosis scale to permit more precise identification of patients who would improve as a result of prefrontal lobotomy. To do so (*a*) the pertinent literature in the field was surveyed to find promising items of predictive significance; (*b*) these items were tried on certain published case histories; (*c*) the items which survived this procedure were combined into a preliminary scale and then used to evaluate, postdictively, 55 lobotomized patients at the VA Hospital, Canandaigua, New York; (*d*) the scale was then cross validated, the stability of the cutoff score was tested, and rater reliability was established for the scale on 60 lobotomized patients at the VA Hospital, Palo Alto, California; and finally (*e*) the scale was revised and improved by essentially repeating steps *c* and *d*. Only the final form of the scale, titled the Palo Alto Lobotomy Prognosis Scale, is described here.

---

[1] From VA Hospital, Palo Alto, California.

[2] Becker selected the items of this scale, tried them out on published case data, developed the original prognosis scale and its manual, collected and analyzed the Canandaigua data, and served as a rater on the Palo Alto study. McFarland was in charge of the collection and analysis of the Palo Alto data, served as a rater at Palo Alto, and prepared the revised scale titled the Palo Alto Lobotomy Prognosis Scale and its manual. Dr. K. P. Jones and Miss Mary A. Hansen served as raters in the Palo Alto study and cooperated in the preparation of a preliminary report of that study. The authors wish to thank B. F. McNeal for his suggestions during the early phases of this study, S. D. Schultz for his assistance in making the improvement ratings, Robert E. Billings for his help in the statistical computations, Mary Ann Swanson for her aid throughout the study, and, finally, the managers of the VA hospitals at Canandaigua, New York, and Palo Alto, California, who made this study possible.

[3] Authority is granted for professional workers to mimeograph, for their own use, the Palo Alto Lobotomy Prognosis Scale data sheets. Copies of the test forms, template, conversion table, and manual have been deposited with the American Documentation Institute. Order Document No. 4468 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Methods and Subjects

Preliminary Prognosis Scale

A preliminary scale of 32 items was constructed from factors found to have some predictive value by Greenblatt *et al.* (2) in their study of 205 lobotomy patients at the Boston Psychopathic Hospital. This source for the initial pool of items[4] was selected because of the comprehensiveness of the variables rated, the relative ease with which the data for rating could be obtained, and the general agreement of the Boston study with other studies. The items included aspects of premorbid personality, history, preoperative psychological functioning, behavior, and psychiatric diagnosis. Three mental status items were converted from ratings to psychological test scores. Other modifications of item definitions were made when necessary to increase clarity.

A trial run of the 32-item prognosis scale was made using data contained in 24 published case histories from the Columbia-Greystone study of cortical ablation (4). Some items were dropped because of their overdependence on the nature of the sample (e.g., religion, age, occupation). Others were modified or eliminated because of difficulty in rating or lack of a significant trend. The results were generally encouraging and 23 items were retained for further study.

Scale-Construction Study

To test the discriminating power of the 23-item prognosis scale and to determine a useful cutting score, it was applied in July, 1953 to the records of 55 male neuropsychiatric patients lobotomized consecutively at the VA Hospital, Canandaigua, New York, between 1948 and 1951. The scale items were rated from data in the clinical folders such as doctor's progress notes, board reports, social service reports, psychological reports, and nurses' notes. Where data were lacking, the total score was prorated on the basis of ratable items.

Criterion improvement ratings. Since the purpose of the study was to postdict improvement or nonimprovement of behavior of lobotomized patients, it was necessary to have a criterion measure which would assess both preoperative and postoperative behavior of these patients and which would yield some quantified measure reflecting the degree of change, if any. Ratings of improvement two to five years after operation were made on the basis of the following scale, borrowed from the Columbia-Greystone study (2, p. 389):

Rating of 5—on most disturbed ward.
Rating of 4—on moderately disturbed ward.
Rating of 3—still institutionalized but on ground privileges.
Rating of 2—at home under supervision.
Rating of 1—at home and working or capable of working.

The improvement measure was the difference between the postoperative and preoperative score plus one bonus point (if appropriate) for each of the following postoperative characteristics:

a. A change from a most disturbed ward to ground parole.
b. A move from any place in the hospital to home.
c. The ability to work at reduced capacity.
d. The ability to work at former capacity.

The maximum possible score was 8.

The improvement criterion had shortcomings, but for the population for which it was designed and used, it appeared to be sensitive enough to allow gross evaluation of improvement.

Malignancy rating. A third measure, a preoperative malignancy estimate, was made to shed further light on the nature of the prognosis scale. In the Columbia-Greystone study (2), it was found that malignancy ratings correlated moderately with improvement. Data obtained in applying the 32-item draft of the prognosis scale to the Greystone sample suggested that this estimate might be thought of as a crude measure of malignancy when malignancy is defined and rated as follows (2):

1. Purely affective psychosis.
2. Schizophrenic features plus considerable affect, little or no deterioration.
3. Schizophrenic features with deterioration, little or no affect, inappropriate affect.

It was tentatively assumed that these concepts crudely reflect an underlying trait continuum.

---

[4] The word "items" is used for convenience only. Actually we are dealing with dichotomized continuous variables in many instances.

Cross-Validation Study

The design of the Palo Alto study was essentially the same as that at Canandaigua except that provision was made to check the reliability of the prognosis and criterion improvement ratings, and to have the improvement ratings made by different judges. Four raters were each assigned 30 records from among the first 60 patients who received lobotomies at the VA Hospital, Palo Alto. The raters included a staff psychiatrist of diplomate status, a staff psychiatric social worker, a staff psychologist, and a clinical psychology trainee. Two judges independently rated each patient on the prognosis and malignancy scales. An averaged rating was used in each case in the statistical computations. A research assistant rated each patient on the improvement scale and a second staff psychologist rerated the first 29 cases for reliability purposes. One brief training session and a preliminary rating manual were used to guide the prognosis ratings.

Description of the Samples

Some population characteristics of the Canandaigua and Palo Alto samples are given in Table 1. From the malignancy ratings, it appears that the Canandaigua sample was somewhat more deteriorated as a group. The differential improvement rates for the two samples is partly accounted for on this basis.

Table 1

Selected Population Characteristics of the Scale Construction and Cross-Validation Samples

Variable	Canandaigua sample	Palo Alto sample
N	55	60
Age in years		
Mean	35.0	39.4
Sigma	9.4	11.3
Range	22–58	23–66
Years from first hospitalization		
Mean	6.2	11.2
Sigma	4.2	8.1
Range	1.5–25	2.2–34.9
Years since operation		
Mean	3.7	3.1
Sigma	.96	1.12
Diagnosis		
Schizophrenic reaction		
Paranoid type	24	23
Catatonic type	13	12
Hebephrenic type	14	18
Mixed type	3	3
Manic-depressive reaction		
Manic (chronic) type	1	0
Depressed type	0	3
Malignancy rating		
1	0	2*
2	19	27*
3	36	31*

* Average of two raters.

Revision of the Scale

On the basis of a chi-square item analysis the 23-item scale was revised. Using the available data from three populations (Greystone, Canandaigua, Palo Alto), 14 of the 23 items were significant beyond the .05 level and were included in the final form, called the Palo Alto Lobotomy Prognosis Scale. Two additional items (sociability prior to illness and undertalkativeness) were also retained, though they did not show statistical significance. Since these items have been consistently found of value in other studies (9), it was felt that with predictive rather than postdictive use of the scale, the data necessary to rate these factors reliably would be available. This assumption will have to be tested later. Items eliminated from the scale were of two kinds: items on which nearly all lobotomy patients scored in the same direction, and items of extremely low reliability. The items of the final scale stated in a prognostically negative direction were as follows:

1. No previous remissions (not home 3 months or more) (1) [5]
2. Irritable prior to illness (1)
3. Not sociable prior to illness (1)
4. Failure to show temporary moderate to marked improvement following convulsive shock therapy (2)
5. Undertalkative speech (1)
6. Incoherent ramblings and "word salad" speech (1)
7. Appearance shabby (1)
8. Presence of auditory hallucinations (2)
9. Absence of awareness of mental disorder (1)
10. Absence of elated, depressed, or anxious affect (2)
11. Presence of flat, apathetic mood (2)

---

[5] Weights are indicated in parentheses. Instructions for rating these items and for determination of prognosis scores are contained in the manual referred to in footnote 3.

12. Disoriented to time, place, or person (1)
13. Wechsler-Bellevue Arithmetic weighted score 6 or under (1)
14. Wechsler-Bellevue Information weighted score 6 or under (1)
15. Wechsler-Bellevue Memory Scale 80 or under (2)
16. The following diagnoses *only* are weighted as indicated:
    - *a*. Involutional psychosis (0)
    - *b*. Manic-depressive, depressed (0)
    - *c*. Manic-depressive, manic (chronic not oscillating) (1)
    - *d*. Schizophrenic, paranoid (1)
    - *e*. Schizophrenic, catatonic (2)
    - *f*. Schizophrenic, mixed (2)
    - *g*. Schizophrenic, hebephrenic (3)

This scale was again scored for both samples and the new cutting score tested.

Results

The 23-Item Prognosis Scale

Reliability. Rater reliability of the 23-item preliminary prognosis scale was found to be .88 on the Palo Alto sample. Rater reliability of the criterion rating was .92 for the first 29 cases in the Palo Alto sample.

Validity data. In computing validity statistics, two cutting scores had to be determined, one on the improvement criterion and one on the prognosis scale. Successful response to lobotomy was defined as improvement to ground privileges or better, which cuts the criterion of improvement between a rating of 1 ("less of a management problem") and 2 ("has some ground privileges"). The effect this definition of successful lobotomy has on the usefulness of the scale will be discussed later. The cutting score on the new prognosis scale was based on a compromise between the best statistical prediction (maximum true positives and true negatives) and an attempt to detect as many as possible that might improve to criterion (minimum false negatives).

If we accept an improvement to ground privileges or better as adequate justification for lobotomy, and use a cutting score determined from the Canandaigua sample by the above procedure, the following validation and cross-validation statistics obtain:

If the 23-item prognosis scale would be employed (using our cutting score) to select patients for lobotomy, (*a*) the percentage of successful operations for the Canandaigua sample would be raised from 22 to 38%. Only two patients who improved to criterion would be overlooked (false negatives), and the number of operations could be reduced by more than half; (*b*) on the Palo Alto sample with the same cutting scores, use of the prognosis scale would raise the percentage of successes from 35 to 58%. Five patients who improved would be overlooked, and again the number of operations could be reduced by more than 50%; (*c*) the percentage of successful predictions is 67% for the Canandaigua sample and 73% for the Palo Alto sample;[6] (*d*) chi squares for each sample are significant beyond the .01 level.

Palo Alto Lobotomy Prognosis Scale

The revised scale has the same interrater reliability (.88) as the original scale even though it is seven items shorter. Distribution data for the validation and cross-validation study are presented in Table 2.

With the same procedure for determining the cutting scores as for the original scale, the following validation data obtain: (*a*) If this scale were used to select patients for lobotomy operations and the cutting score (as indicated in Table 2) derived from the Canandaigua data were used, the percentage of successful lobotomies in the Canandaigua sample would be raised from 22% to 35%. Only two patients who improved to criterion would be overlooked (false negatives) and the number of operations could be reduced by 50%. (*b*) On the Palo Alto sample if this scale were employed using the cutting score from the Canandaigua sample to select patients for lobotomy, the percentage of successes would be raised from 35% to 59%. Two patients who improved would be overlooked, and the number of operations could be more than halved. (*c*) The percentage of successful predictions is 65 for the Canandaigua sample and 73 for the Palo Alto sample.[7] (*d*)

---

[6] With a cutting score determined from the Canandaigua sample to give the best statistical prediction, the percentage of successful predictions is 73% for Canandiagua and 78% for Palo Alto.

[7] With a cutting score (between 10 and 11) determined from the Canandaigua sample to give the best statistical prediction, the percentage of successful predictions is 78% for both samples.

Table 2

Distribution of Cases for the Relationship of the Prognosis Score to Improvement

Prognosis score*	Canandaigua improvement rating		Palo Alto improvement rating	
	0–1	2–8	0–1	2–8
2–6	2	2	0	6
7–10	5	5	3	5
11–14†	10	3	10	8
15–18	16	1	16	1
19–23	10	1	10	1
Total *N*'s	43	12	39	21

* Lower prognosis scores and higher improvement ratings are in the favorable direction.

† The cutting score on the prognosis scale falls between 14 and 15.

Chi squares based on the cutting scores indicated in Table 2 are significant beyond the .01 level for each sample.

Malignancy Ratings

The interrater reliability of the malignancy ratings was .70 for the Palo Alto sample. As was suggested from applying the preliminary scale to the Greystone sample, the prognosis scale shows a high relationship to malignancy. In the Canandaigua sample the biserial *r* was .79, and in the Palo Alto sample the product-moment *r* was .83.[8] So high a relationship indicates that the prognosis scale is primarily a measure of the degree or severity of illness. If it were not for its coarse nature and lower reliability, the malignancy rating by itself would be an adequate prognostic device. It correlates with improvement just as highly (Pearson *r*, .50) as the prognosis scale (Pearson *r*, .48) for the Palo Alto sample. However the prognosis scale has the advantage of finer differentiation leading to better prediction.

Implications

Malignancy and Improvement

Analysis of the relations between the prognosis scale scores and malignancy ratings strongly supports the idea that patients who improve as a result of prefrontal lobotomy are less mentally ill than those patients who do not profit from lobotomy. This idea is not a new one. Stotsky's (7) study of remitting and nonremitting schizophrenics shows that remitters are more like severe neurotics on the Rorschach than nonremitters. The study of process and reactive schizophrenics by Kantor, Wallner, and Winder (3) leads to a similar conclusion. Wittman's work (8) with the Elgin Prognosis Scale also supports this thesis. In the neurotic range of severity, Barron (1) has demonstrated that the neurotics who profit most from psychotherapy are better adjusted prior to treatment. The apparent truth of this thesis has the following implications for further work in this area. First, the search for prognostic indicators should be more actively centered on variables which correlate with severity of illness. Items in this lobotomy prognosis scale suggest many areas for more intensive study. Second, it becomes essential, particularly with small samples, that studies of the differential effects of various therapies be controlled for severity of illness. The prognosis scale discussed in this paper might be useful for this purpose with some severely ill populations.

Uses and Limitations of the Scale

a. Present validity data are limited to male veterans in VA hospitals referred on some basis other than their scores on this scale as lobotomy candidates.

b. Our definition of "success" on the criterion improvement rating has a definite effect on the utility of the prognosis scale as a selection device. We accept as our definition of success an improvement to ground privileges or better (two or more points on the improvement scale). If one chooses the less stringent criterion of better ward management (an improvement of one point), the prognosis scale has little value in selection, since most patients under present selection methods improve this much. However, if the patient has a negative prognosis score in terms of our cutoff point, the scale would indicate that better management would be the most that could be expected. Then the decision to operate or not, under this prediction of limited improvement, could be weighed against such factors as possible defects the lobotomy operation creates, the surgical risks involved, and the cost of the procedure.

---

[8] The fact that the correlation between malignancy ratings and prognosis scores on the Palo Alto sample is higher than the reliabilities would allow is possibly a function of the increased reliability obtained from averaging the two ratings. Chance factors may also enter.

c. A favorable score on the scale does not insure success from lobotomy. However, the probability of success, accepting our criterion of improvement, can be stated to parents or relatives who must approve the operation. It must be pointed out that the revised prognosis scale, from a statistical standpoint, rules out lobotomy operation with greater certainty than it rules in the procedure. Roughly speaking, a score above the cutoff indicates with the odds of 13 to 1 that the patient can be expected to show *little or no* improvement whereas scores below the cutoff indicate with odds of only 1 to 1 that a patient should demonstrate *marked* improvement.

d. A final caution in the use of the scale involves the fact that all of the present data on the utility of the scale are based on postdictive studies. Though there is no reason to believe the scale will not be useful in prediction, it has not yet been demonstrated. The fact that we can identify with greater certainty those patients on whom operation should *not* be performed suggests that other types of psychological and/or physiological data are needed to identify with greater accuracy those patients who *will improve* as a result of lobotomy.

Summary

A series of studies was undertaken to develop and evaluate a lobotomy prognosis scale. Using historical, behavioral, and mental status material found to have some prognostic significance by Greenblatt *et al.* (2), we developed a preliminary 32-item scale. Modifications of the scale were made on the basis of data obtained in applying it to 24 published case histories from the Columbia-Greystone study of cortical ablation. The resulting 23-item scale was then validated against the records of 55 lobotomy patients at the VA Hospital, Canandaigua, New York, and cross-validated on 60 lobotomy patients at the Palo Alto VA Hospital. Evaluation of improvement was made at the time of the study which was from two to five years after operation. Finally the scale was revised on the basis of a chi-square item analysis of the data from the three samples (Greystone, Canandaigua, and Palo Alto).

The results indicated that the prognosis scale predicts to a practical degree both those patients who are likely to improve to ground privileges or better and those patients who demonstrate little or no improvement after lobotomy. The high two-rater reliability of the scale, and the cross-validation stability of the cutting score and percentage of successful predictions suggested that considerable confidence may be placed in the repeatability of the findings. Cautions and limitations are discussed in the text.

In an attempt to clarify the underlying nature of the prognosis scale, ratings of malignancy were correlated with the prognosis scores. Correlations of about .80 found in three samples force a strong conclusion that the prognosis scale is a refined rating of severity of illness.

Received August 19, 1954.

References

1. Barron, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 235–241.
2. Greenblatt, M., Arnot, R., & Solomon, H. C. (Eds.) *Studies in lobotomy*. New York: Grune & Stratton, 1950.
3. Kantor, R. E., Wallner, J. M., & Winder, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.
4. Mettler, F. A. (Ed.) *Selective partial ablation of the frontal cortex*. New York: Hoeber, 1949.
5. Partridge, M. *Pre-frontal leucotomy*. Springfield, Ill.: Charles C Thomas, 1950.
6. Scherer, I. W. Prognosis and psychological scores in electroconvulsive therapy, psychosurgery, and spontaneous remissions. *Amer. J. Psychiat.*, 1951, **107**, 926–931.
7. Stotsky, B. A. A comparison of remitting and non-remitting schizophrenics on psychological tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 489–496.
8. Wittman, Phyllis. A scale for measuring prognosis in schizophrenic patients. *Elgin Pap.*, 1941, **4**, 20–33.
9. Wittman, Phyllis. Diagnostic and prognostic significance of the shut-in personality type as a prodromal factor in schizophrenia. *J. clin. Psychol.*, 1948, **4**, 211–214.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Psychosomatic Patients and Other Physically Ill Persons:
A Comparative Study[1]

Sheldon E. Waxenberg
The Mount Sinai Hospital, New York, N. Y.[2]

Psychologists who work in general hospital settings frequently have occasion to evaluate patients' responses in the specific light of psychosomatic theories. Awareness that these theories have not been subjected to rigorous experimental confirmation causes certain persistent questions to arise, especially if some patients are regarded as psychosomatic cases while others are thought of simply as physically ill people. Are patients with diseases such as asthma and ulcerative colitis, which are widely referred to as psychosomatic because they appear to have mixed psychological and physiological etiology, actually more like psychiatric patients than like patients with ordinary somatic illnesses? Are they more likely to suffer from depressed moods? Do they present characteristically distorted movement and color balances and give more bony anatomy responses on the Rorschach? Other questions also arise, concerned with whether or not the differences in psychodynamics hypothesized for the distinctive psychosomatic maladies evidence themselves in projective test protocols. For example, are oral objects and processes more prominent in the records of asthmatics, and anal objects and processes more notable in the records of colitis sufferers? This paper constitutes an attempt at finding answers to these and other questions arising out of work with a diverse hospital population.

Experimental Subjects and Procedures

In order to carry out meaningful experimental comparisons of patients with psychosomatic diseases, it is essential to choose disease entities which are widely conceived of as within the psychosomatic category and which have distinctive systemic and dynamic features. Asthma and ulcerative colitis qualify in both these respects. For purposes of broader comparison, a group of subjects with a disease definitely not currently thought of as psychosomatic was chosen as an experimental control. Tumor patients appeared to qualify in this respect, for the etiology of malignant tumors has not been directly linked with psychological mechanisms or stress. Furthermore, patients with tumors are relatively as ill as the subjects in the other groups.

Subjects

The subjects, 20 in each group, all women and all 23 to 46 years of age, were clinic and ward patients of one general hospital. The median age of the asthma group was 36.5, of the colitis group 32.5, and of the tumor group 40.8 years. Because of a dearth of suitable colitis subjects in the clinics and wards, the members of a hospital social club for people who have had ileostomies performed to control ulcerative colitis were included in the colitis group. Only patients found to be currently at least occasionally subject to asthma attacks were included in the asthma group. The colitis subjects had all been hospitalized at least once for ulcerative colitis. All the tumor patients had been definitively diagnosed as suffering from malignant breast, uterine, abdominal, or lymphatic system tumors. No one was included whose hearing or sight was severely impaired or who was pregnant, crippled, or receiving ACTH or cortisone. Nor was anyone included who in addition to one of the three selected diseases had a medical history of any other crucial disease.[3]

---

[1] The data were collected at Mount Sinai Hospital, New York City, with the permission of the Director of Medical Research, Dr. Alexander B. Gutman, while the writer was a member of the psychology staff of Dr. Fred Brown. The persons at the hospital who expedited and encouraged this research are, regrettably, too numerous for individual mention.

The present paper is part of a dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Teachers College, Columbia University, 1954. The writer wishes to express his thanks to Drs. Edward Joseph Shoben, Jr., Laurance F. Shaffer, Helen M. Walker, and M. Ralph Kaufman for their invaluable stimulation and advice.

[2] Now at Kings County Hospital, Brooklyn, New York.

Psychological Tests Used

The battery of five tests used in this study, which, except for the omission of a lengthy intelligence scale, resembles a conventional diagnostic battery, comprised the following:

The Bender-Gestalt Test. The Bender records were scored by the experimenter in accordance with Pascal and Suttell's standardization (5) after prescribed practice with their method had resulted in a product-moment correlation of .89 with Pascal's scores for the 20 sample records in his manual. Pascal reports a correlation of .90 between his own and his collaborator's scorings of 120 records.

The H-T-P Test. Drawings of house, tree, and person were supplemented by another drawing. After the first human figure was finished, the subject was requested to draw a figure of the opposite sex.

The Rorschach test. The Rorschach records were scored in accordance with Klopfer's method (4). The proportion of agreements of scoring judgments made by the experimenter and by another psychologist[4] on 20 random records ranged from .59 to .93 for explicit agreements on individual color and movement determinants. When implicit agreements —instances in which neither scorer considered a response to involve a certain determinant—were included in the computations, the proportion of agreements ranged from .71 to .98 on individual determinants. The human and animal movement responses were judged to fall into one of three categories—active, static, or passive—in accordance with a classification scheme based in large measure on a discussion of movement responses by Phillips and Smith (6, pp. 75–77). Scorer reliability on these classifications ranged from .76 to 1.00 on explicit judgments and from .82 to 1.00 when implicit judgments were also taken into account. The Rorschach responses were also judged for oral, anal, and bony anatomy content. A response was considered oral if it involved (*a*) foods or beverages, (*b*) culinary tools or procedures, (*c*) body parts related to food intake, (*d*) oral activities such as biting, sucking, and licking, or (*e*) states of hunger or thirst. A response was considered anal if it involved (*a*) excretive functions or appliances, (*b*) fecal matter, filth, dirt, the implication of a morbid or disgusting textural quality or viscosity, or terms such as "oozing," "squishing," etc., or (*c*) posterior anatomy, either internal or external. A percept was considered a bony anatomy response if it involved an anatomical part, either animal or human, with a specifically bony structure. Scorer reliability on these categories ranged from .71 to .81 for explicit judgments and from .98 to .99 for all judgments.

Word associations. Rapaport's revised list of 60 word association stimuli was used (7, p. 84). The responses were judged by the same criteria as those for the oral and anal Rorschach content, and the number of each kind of response was totaled for each subject. Twenty random records were rescored by another judge. The two scorers agreed on .85 of the oral and .82 of the anal responses. With implicit agreements included, the over-all agreement was .97.

The Thematic Apperception Test. Ten TAT cards (1, 2, 3BM, 4, 8BM, 12M, 13MF, 15, 18GF, 16), always in this order, were used. Each story was rated for emotional tone in accordance with a scale devised by Eron (2, Appendix A). An algebraic sum was computed for the ratings of the ten stories of each subject. The resultant score for every one of the 60 subjects was negatively signed, indicating a weighting toward sad stories, which also prevailed in Eron's studies. Because of this uniformity, the signs were dropped and a set of mood scores was obtained which ranged from 1, indicating a neutral tone, to 14, reflecting a rather deep dysphoria. Twenty random records were rescored by another judge. The product-moment correlation of the two sets of ratings was .82. Eron obtained correlations of .86 between raters in his own studies (2, 3).

Hypotheses and Results

Certain hypotheses were implied in the questions raised in the introductory paragraph. These and other hypotheses suggested by theories and intuitions about personality dynamics which are applied and found empirically useful in clinical work are here stated explicitly and tested by statistical methods.

---

[3] The subjects were screened for an acceptable level of verbal fluency. Some differences were found in vocabulary scores, educational attainment, and ethnic composition of the groups, but these were investigated wherever possible and found not to affect critically the major purposes of the study.

[4] Barbara R. Waxenberg independently rescored a portion of the records.

Psychiatric Patient Status on the Bender-Gestalt Test

Pascal's scoring method for the Bender-Gestalt (5), which differentiates psychiatric clinic and hospital patients from nonpatients, was used in this study to test the hypothesis that the behavior of a person suffering from a psychosomatic disease is more likely to resemble the behavior of a psychiatric patient than is the behavior of an individual with a disease which is not spoken of as psychosomatic. All subjects were dichotomized according to their standard scores, with a score of 67, the point at or beyond which only 5 in 100 normal persons' scores would be expected to fall, serving as a cutoff point. Low scores which placed them in the nonpsychiatric category were obtained by 7 asthma, 13 colitis, and 5 tumor subjects. The balance of the 20 subjects in each group fell in the class resembling psychiatric patients. When the asthma and colitis groups were compared by chi square, a significant difference did not appear.[5] Since they could then be regarded as drawn from the same population, they were combined into a composite psychosomatic group for comparison with the control tumor group. The difference between these groups was not significant either. Hence, it cannot be concluded that people with psychosomatic diseases differ from other ill people in respect to psychiatric status.

Mood Scores on the Thematic Apperception Test

Eron's rating scale for emotional tone of TAT responses was designed to discriminate between stories expressive of aspirations and pleasant expectations and stories of fearsome, frustrating, and violent events associated with feelings of guilt, regret, and resignation. Psychosomatic patients are considered to be immobilized by conflicting needs and motivations which produce dysphoria. The hypothesis that such patients are more subject to dysphoric moods than are other sick people was tested by means of dichotomization of the subjects on the basis of TAT mood scores falling above or below the median of 9.4. Scores indicating greater dysphoria than the median score were obtained by 10 asthma, 11 colitis, and 8 tumor patients. Group comparisons by means of chi square, similar to those described in the preceding section, produced no evidence to indicate that dysphoria is more commonly a concomitant of psychosomatic illness than of other physical illness.

Sequence and Size of Male and Female Figure Drawings

In psychosomatic formulations, asthmatics are regarded as passive, dependent, and very much identified with supportive maternal figures, while women afflicted with ulcerative colitis are regarded as assertive and dominating and often career oriented. The personal projections obtained in human figure drawings were used to test these psychosexual differentiations. It was hypothesized that women asthmatics, in carrying out the figure-drawing task, draw the female figure before the male figure and tend to draw the female larger in vertical dimension than the male figure, while women with ulcerative colitis reverse the process, drawing the male first and larger. The female was drawn first by 13 asthma, 16 colitis, and 14 tumor subjects. Pairs of figures were regarded as similar in size whenever the difference in their vertical dimensions, measured to the nearest half centimeter, was less than 1.5 centimeters. In only 39 cases were there differences this large or larger. The female figure was drawn larger by 11 asthmatics as against 4 who drew the male larger, by 6 colitis patients as against 6 who drew the male larger, and by 10 tumor patients as against 2 who drew the male larger. When the asthma and colitis groups and then the combined psychosomatic and control groups were compared on each of these variables by means of chi square, no significant differences emerged.

---

[5] Yates' correction for continuity was applied whenever a least expected frequency was less than 10, except where its application would operate to increase chi square. The .05 level of confidence is accepted as defining significance throughout this study.

Hypotheses Concerned with Rorschach Measures

Consideration was given to the factor of total productivity before counts of Rorschach determinants were used for testing experimental hypotheses. The number of responses in each of the 60 Rorschach records ranged from 7 to 43. The median fell between 15 and 16. Fewer than the median number of responses were given by 14 asthma, 10 colitis, and 9 tumor patients. Chi squares offered no basis for doubting the homogeneity of these groups with respect to total responsiveness. However, since the number of responses in an individual record is correlated to some extent with each of the major determinants, it is advisable wherever practicable to examine each hypothesis in such a way as to impose a degree of control on productivity. The method adopted here is to test a hypothesis first with only those records containing fewer than the median number of responses, then repeat the test with only those records having more than the median number of responses, and finally repeat the test using all the records.

Coarctation. When both the total of human movement responses and the weighted color sum fail to exceed one, a record is considered to be coarctated. Observation in psychiatric clinics devoted to treatment of psychosomatic cases suggests a prevalence of coarctation in Rorschach productions of this group. The question arises, however, whether it is primarily a matter of the patient's being physically ill in general or a matter of his having a disease regarded as psychosomatic. It was here hypothesized that Rorschach records of psychosomatic patients are more likely to show coarctation than are the records of other physically ill persons. Coarctation was found in the shorter records of five asthma, three colitis, and four tumor patients and in the longer record of one tumor patient. Chi squares revealed no significant differences in comparisons of the asthma and colitis groups or of the combined psychosomatic and control groups, when shorter or longer or all records together were considered.

Human movement and color response patterns. Instead of the coarctative manifestations described above, a somewhat different patterning of color and human movement responses is considered by Phillips and Smith (6) to typify the psychosomatic patient. Their assertions are embodied in the hypothesis that psychosomatic patients typically give no more than one human movement response but obtain weighted color sums of between three and five. The number of subjects in this study who show precisely this pattern is very small. Three members of the control group produced the pattern in longer records; one asthma and two colitis patients did so in shorter records. For further testing of the hypothesis, the two components of the pattern were investigated separately. Of those producing shorter records, 13 asthma, 8 colitis, and 9 tumor patients obtained color sums of less than three; while 1, 2, and 0 patients, respectively, obtained color sums of between three and five. Of those producing longer records, 5 asthma, 1 colitis, and 5 tumor patients obtained color sums of less than three; while 1, 7, and 5, respectively, obtained color sums of between three and five. The total color sums were 22.0, 54.5, and 48.5 for the respective groups. Two longer colitis records and one longer tumor record show color sums between 5.5 and 7. These three records were excluded from the tabulations; however, had they been included, the same statistical results would have been obtained. When the shorter record asthma and colitis categories were compared and when they were combined for comparison with the control group, no significant chi squares were obtained. When the longer records, or all the records, of the asthma and colitis subjects were compared, however, chi squares significant at better than the .05 level of confidence were obtained, indicating that colitis subjects are more likely to give numerous color responses than are asthmatics. Comparisons of the tumor group with the asthma group and with the colitis group, each taken separately since lack of homogeneity forestalls combining them, offered no evidence of differences between the individual psychosomatic groups and the control group.

Now, what may be said of the other half of the pattern? Of those producing shorter records, 8 asthma, 7 colitis, and 7 tumor patients gave one M response or no M at all;

while 6, 3, and 2 patients, respectively, gave more than one *M* response. Of those producing longer records, 5 tumor patients gave no more than one *M* response; while 6, 10, and 6 patients, respectively, gave more than one *M* response. The respective groups produced totals of 51, 47, and 28 *M* responses. When comparisons similar to those described above were made on the human movement variable, only one significant difference was found, that between the longer record subjects of the composite psychosomatic group and of the control group. However, the direction of the difference was opposite to that hypothesized. About half of the 11 control group subjects who gave longer protocols fall in the 0–1 *M* category, whereas not one of the 16 psychosomatic subjects who gave longer records had such a small number of *M* responses. From all these findings, there would appear to be very little reason for believing that the color-movement pattern proposed by Phillips and Smith is characteristic of psychosomatic records.

Active and passive ideation in movement responses. Human and animal movement responses on the Rorschach were separately differentiated as being active, static, or passive for the purpose of testing the hypothesis that asthma patients predominantly project basic passivity and dependency needs while colitis patients reflect rather strong strivings and aggressive drives in their fantasy productions. After rather complex categorizations were made, chi-square tests on the resultant data brought to light no significant differences between the groups.

Static human movement responses. Static human movement responses reflecting inhibition of hostile activity and resultant tension in situations of conflict between activity and passivity indicate psychosomatic symptomatology according to Phillips and Smith (6, p. 77). The median number of such responses per record was between 0 and 1 for the subjects of this study. The records of two asthma, eight colitis, and three tumor subjects contained at least one static *M* response. When chi-square tests were made on the basis of the presence or complete absence of static *M*'s, no substantiation was obtained for the hypothesis that psychosomatic patients are more likely to produce static *M*'s than are other physically ill people.

Bony anatomy. In an extensive review of types of Rorschach content, Phillips and Smith (6, pp. 124–127) declare that bony anatomy content is an index of the extent to which psychosomatic patients are engaged in inhibiting the expression of hostile impulses which are ultimately expressed through somatization. The hypothesis that psychosomatic subjects are more likely to produce bony anatomy responses than are other sick persons was tested by classifying the subjects according to whether they gave no bony anatomy responses or at least one bony anatomy response, the median lying between these limits. One or more such responses were given by 11 asthma, 9 colitis, and 11 tumor patients. The responses total 24, 11, and 32 for the respective groups. Chi squares gave no evidence of significant differences.

Oral and anal Rorschach content. It is hypothesized that asthmatics, considered to be conflictually absorbed with problems of succorance and maternal protection, fill their Rorschachs with inordinate amounts of oral content, while colitis patients produce an unusual amount of anal content, either as an expression of deep, hidden preoccupations and motivations or as a reflection of pressing concern with the problem of fecal elimination. The median number of oral responses, and also of anal responses, for all subjects was between 0 and 1. One or more oral responses were given by 9 asthma, 10 colitis, and 11 tumor patients, for group totals of 12, 21, and 20, respectively. One or more anal responses were given by 1 asthma, 1 colitis, and 5 tumor patients, for group totals of 1, 3, and 12. Chi-square tests revealed no significant differences.

Oral and Anal Responses to the Word Association Test

The hypothesis involving oral and anal responses was also investigated by means of a word association list (7, p. 84) in which the total number of such responses bulks much larger and might be expected to afford a better differentiation of the groups. Oral responses ranged from 3 to 13 per record, with

Bureau Edl. Psy. Research
Dated
Accs No

the median between 8 and 9. More than 8 oral responses were given by 10 asthma, 11 colitis, and 11 tumor patients. Group totals were 177, 174, and 172, respectively. The number of anal responses ranged from 0 to 5 per record, with the median between 2 and 3. More than 2 anal responses were given by 7 asthma, 10 colitis, and 9 tumor patients. Group totals were 42, 52, and 52. Chi-square tests failed to sustain the hypothesis derived from psychodynamic and symbolistic theorizing about the expression of fixated pregenital oral and anal drives.

Discussion

The primary purpose of this study was to lend the support of experimental evidence to clinical impressions about people with psychosomatic diseases. It is widely held that they constitute a unique group, different from normal, healthy persons as well as from persons physically ill in the ordinary sense. Moreover, the psychodynamic forces at work in various such diseases are said to differ in particular ways. By means of a conventional test battery, these assumptions were investigated, without any evidence for their soundness being adduced. Such consistently negative findings cast doubt upon the basic formulations which set the direction for psychosomatic practice, therapeutic as well as diagnostic. The selection of a presumedly normal but physically ill, rather than a normal and physically healthy, control group placed the focus on two polarities of physical illness, psychosomatic and nonpsychosomatic. An attempt at delineating the differences between the two has rendered questionable the theoretical differentiations which have been posited in the literature. Dunbar, almost twenty years ago, implied the inevitability of a resynthesis in thinking about the two when she wrote, "When medicine has apprehended the psychosomatic problem and assimilated it, the adjective will be obsolete: all medicine will be psychosomatic" (1, p. xx).

A question about the optimum utilization of the obtained data deserves attention. Might not global clinical evaluations of such phenomena as moods, identifications, preoccupations, and so forth, lead to different or at least more comprehensive conclusions than did the disparate, atomistic measures used in this study? An experimental design requiring a number of experienced psychological clinicians to rate certain selected crucial variables on the basis of all the material in each subject's complete protocol of responses might make fuller use of the data. Such an alternative approach, which is being planned, might afford a more convincing and thoroughgoing examination of the psychosomatic theories. In any event, it is clear from the foregoing findings and discussion that all major psychosomatic formulations need to be subjected to rigorous and searching analysis by a variety of techniques.

Summary

To investigate certain hypotheses about psychosomatic disease, experimental groups of 20 asthmatic women and 20 women with histories of ulcerative colitis and a nonpsychosomatic control group of 20 women being treated for malignant tumors, all between 23 and 46 years of age and screened to eliminate patients with multiple diseases, were given a test battery including the Bender-Gestalt, figure drawings, Rorschach, word associations, and ten TAT cards.

Bender-Gestalt scores gave no evidence that psychosomatic patients differ from other sick people in psychiatric status. TAT mood scores gave no indication that psychosomatic patients are more subject to dysphoria than are other physically ill persons. Rorschach records revealed no significant evidence that psychosomatic patients are more likely to (*a*) give coarctated records, (*b*) give fewer human movement responses, (*c*) obtain higher weighted color sums, (*d*) produce static human movement responses, or (*e*) produce bony anatomy responses than are others physically as ill as they are. Rorschach movement responses did not reveal significant tendencies for asthmatics to project passivity and dependency needs or for colitis patients to project strivings and aggressive drives in their fantasy productions. No evidence was derived from word associations or Rorschach content to indicate that asthmatics are inordinately concerned with oral satisfactions and deprivations or that colitis sufferers are inordinately preoccupied with excre-

tory functions and their indirect, symbolic counterparts. Neither the sex sequence of the two human figure drawings nor their relative size offered any evidence of differences in the psychosexual identifications of women with asthma and women with ulcerative colitis. The consistently negative trend of the findings and implications for theory and research were discussed.

Received September 18, 1954.

References

1. Dunbar, Helen Flanders. *Emotions and bodily changes.* (2nd Ed.) New York: Columbia Univer. Press, 1938.
2. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, 64, No. 9 (Whole No. 315).
3. Eron, L. D. Responses of women to the Thematic Apperception Test. *J. consult. Psychol.*, 1953, 17, 269–282.
4. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
5. Pascal, G. R., & Suttell, Barbara J. *The Bender-Gestalt test: quantification and validity for adults.* New York: Grune & Stratton, 1951.
6. Phillips, L., & Smith, J. G. *Rorschach interpretation: advanced technique.* New York: Grune & Stratton, 1953.
7. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Vol. 2. Chicago: Year Book Publishers, 1946.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Parental Figures in Sentence Completion Test, in TAT, and in Therapeutic Interviews

Mortimer M. Meyer and Ruth S. Tolman
Mental Hygiene Clinic, VA Regional Office, Los Angeles

A follow-up study to that reported in "Correspondence Between Attitudes and Images of Parental Figures in TAT Stories and in Therapeutic Interviews" (1) was carried out with the Forer Sentence Completion Test.

Twenty of the 50 patients used in the TAT study had taken the Sentence Completion Test. The 11 items (9, 16, 33, 35, 60, 70, 76, 88, 94, 96, 99) referring to "mother" and "father" were examined, and parental attitudes and images discovered there were indicated on the same check sheets used in the study of the TAT and of therapeutic interviews.

Analysis of the data by the same method gave the following results:

1. When the Sentence Completion Test and the TAT were compared in regard to fathers' attitudes, six of the 20 had one item in common and one had two in common. Of the seven records, four could be interpreted as challenging the null hypothesis at a level of confidence between 1 and 10 per cent.
2. When the Sentence Completion Test and therapeutic interviews were compared in regard to fathers' attitudes, only five had any item in common and none more than one. Of these five only three could be interpreted as challenging the null hypothesis at a level of confidence between 5 and 10 per cent.
3. When the Sentence Completion Test and the TAT were compared in regard to mothers' attitudes, only two had any item in common, neither more than one, and neither could be interpreted as significant.
4. When the Sentence Completion Test and therapy sessions were compared in regard to mothers' attitudes, five had one item in common, none more than one, and only one challenged the null hypothesis at a level of confidence between 5 and 10 per cent.
5. In regard to fathers' images, comparison of the Sentence Completion Test with the TAT and with therapy sessions shows only one item in common in each comparison. Neither is significant.
6. In regard to mothers' images, comparison of the Sentence Completion Test with the TAT and with therapy sessions shows two items in common in each comparison. None of these is significant.

These results suggest that the kinds of attitudes and images of parents which are given in the Sentence Completion Test furnish no basis of prediction of those found in either TAT stories or in therapy sessions.

Received August 19, 1954.

Reference

1. Meyer, M. M., & Tolman, Ruth S. Correspondence between attitudes and images of parental figures in TAT stories and in therapeutic interviews. *J. consult. Psychol.*, 1955, **19**, 79–82.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Some Standardized Scales for Disorganization in Schizophrenic Thinking[1]

William A. Hunt
Northwestern University

and Franklyn N. Arnhoff
University of Nebraska College of Medicine

In descriptive psychopathology we often teach by illustration. Thus, in speaking of confusion in schizophrenic thinking, we might illustrate it by quoting at random several examples of schizophrenic responses to vocabulary items on a standard vocabulary test. Such clinical materials are rarely presented in the order of the severity of pathology exhibited, but usually are put before the student in random, unorganized fashion. It would be more meaningful, pedagogically, if they could be presented in some kind of orderly scale. The difficulty of presenting clinical material in this fashion is that we have no "standardized" materials from which to draw. Pathological material is rarely ordered or scaled in this fashion, perhaps, because of the assumed unreliability of the clinical judgments which must be the basis for such a scaling approach.

In a previous study (1) in which 22 experienced clinical psychologists rated, on an 11-point scale for disorganization of thinking, standard vocabulary item responses made by schizophrenic patients, we found indications that there might be sufficient agreement among clinicians on such items for scaling purposes. We, therefore, decided to try constructing such scales, limiting ourselves to pathology as manifested on verbal test responses because of the ready availability of the material and the ease with which it can be reproduced without photographic or phonographic recording.

Procedure

In the previous study (1), 222 schizophrenic responses to Wechsler-Bellevue and Terman-Binet Vocabulary items, judged by three trained clinicians as covering all possible values of confusion in thinking in such responses, were rated by 22 clinical psychologists of at least four years' professional experience using an 11-point scale for degree of disorganization in thinking. On the basis of the means and standard deviations of the raters' judgments for these items, we selected 50 of them for the present study. These items were evenly distributed over the 11-point scale used and seemed to represent an adequate sampling of the original universe of items.

In an attempt to extend the range of our materials we also selected 50 responses to Comprehension items on the Wechsler-Bellevue Intelligence Scale. These were chosen from the same original test protocols from which the vocabulary items were drawn, and were judged as being representative of the complete range of pathology exhibited. Other types of test materials such as Information and Similarities were considered for use but were discarded after inspection of a large number of such responses revealed that they did not offer a suitable range of pathology for the construction of scales with more than a minimum of scale points.

For the present study, it was felt more desirable to have the judges rate on a 7- rather than an 11-point scale. Furthermore, the instructions for the subjects in the present study were modified, as the previously used instructions were felt to be too vague. Indications from the data and discussions with the subjects of our first study indicated that other criteria than mere severity of pathology were being used. Prognosis, therapeutic indications, etc. were all being used, confounding the variable under study and lowering the reliability of the judgments. We, therefore, wrote very careful, explicit instructions, defining the task requested of the judges. This point is of considerable importance because of the reliability of the results of the present study, which is much higher than that previously discovered in our own work (1, 4) and that of others (2, 6). The instructions follow:

---

[1] This study is part of a larger project continuing under ONR contract 7 onr–450(11) with Northwestern University. The opinions expressed, however, are those of the individual authors and do not represent the opinions or policy of the Naval service.

> We are going to present you with a number of schizophrenic test responses to items taken from the Wechsler-Bellevue and Terman intelligence scales. One of the ways in which the pathology of schizophrenia may express itself is through disordered thinking which results in atypical, deviant, or "abnormal" responses to the items on such a test. The qualitative interpretation by the clinician of such test responses is one of the bases upon which he may base a clinical or diagnostic interpretation. All of these responses were given by schizophrenics. None of the cases were complicated by mental deficiency or known organic conditions. The extent of the pathology exhibited however is not uniform. In some of the responses it is minimal and in others it is extreme.
>
> We are going to ask you to rate these responses on a 7-point scale according to the severity of the pathology exhibited in the response, with the low end of the scale representing minimal pathology and the high end of the scale representing maximal pathology. In making these ratings we are asking you to concentrate upon the severity of the pathology exhibited in the response itself. We are not interested in such things as prognosis, chronicity, therapeutic indications, etc. In essence, what we are asking you to do is to judge how "schizophrenic" each response is.
>
> Our basic assumption is that the nature of such responses is one of the dimensions upon which a diagnosis of schizophrenia is made. We are assuming that some kind of a continuum exists and that these items could be placed upon it, i.e., the pathology exhibited in some of these responses may be such that in and of itself it would not lead you to think that the persons responding were schizophrenic. Others of them might cause you to suspect schizophrenia. In still others, the pathology exhibited may be so extreme as to make you reasonably sure that the patient is schizophrenic. We realize that no actual diagnosis would ever be made on the basis of such responses alone. In this sense our situation might be called somewhat artificial. We are not asking you to make a diagnosis of schizophrenia on a patient, however. *We are merely asking you to rate how schizophrenic each of these responses is.*
>
> We are doing this because we wish to see whether or not there will be sufficient agreement among clinicians so that we may establish scales representing the typical responses of schizophrenics. The obvious analogy would be with a handwriting scale or a colorimeter in chemistry.
>
> Our hope is that even if such scales would not be applicable in the diagnostic situation as an aid in evaluating the nature of actual test responses or making a clinical interpretation of actual test results, they nevertheless would be very helpful for teaching purposes. Thus, in introducing students to the pathology exhibited in schizophrenic thinking, instead of introducing random illustrations of peculiar responses given by schizophrenics, it would be of real value to present them with an orderly series of test responses, scaled on a 7-point continuum based on the empirically ascertained agreement of a group of trained clinicians. Our goal is to find whether or not such scales are possible and, if so, to construct them. Your judgments on this experiment will be the basis for our work. We thank you for your cooperation and hope that you will agree with us that such work is valuable in attaining the long range goals of our mutual profession.

The 50 Vocabulary responses and the 50 Comprehension responses were then presented to 16 clinical psychologists working in the Chicago area. Each one possessed the Ph.D. in clinical psychology and a minimum of four years on-the-job clinical experience. The data were mimeographed and presented so that Vocabulary items were rated first and Comprehension items second, with the same sheet of instructions serving for both. Due to the relatively small N, counterbalancing the order of presentation was not attempted.

In order to check the reliability of our judges and materials, the task was repeated three months later. The subjects were initially unaware of our intent to request a retest. The group repeat reliability was .97 for Vocabulary and .96 for Comprehension, using a Pearson r. The individual test-retest coefficients using the Pearson r for the individual judges ranged from .65 to .92 for Vocabulary and from .68 to .90 for Comprehension. The agreement among the clinicians was measured by correlating each individual's judgments on each group of 50 items with the average of the group of judges, using a Pearson *r*. These ranged from .73 to .92 for Vocabulary and from .64 to .88 for Comprehension on the first presentation, and from .69 to .92 for Vocabulary and from .66 to .86 for Comprehension on the repeat performance. These reliabilities were considered sufficiently high

Table 1

Vocabulary Scale

Scale Point	Response	Mean	*SD*
1	Gamble—To take a chance, a risk	1.00	0.00
	Seclude—To go away and be alone, to seclude oneself	1.50	0.63
	Donkey—A type of four-legged animal	1.50	0.52
2	Gown—Garment you wear for lounging	1.75	0.93
	Shrewd—Careful in a sneaky, clever way	2.19	0.75
	Nail—A bit of metal used to pound on	2.37	0.81
3	Plural—Means plus another	2.94	0.93
	Join—Has to do with organization	2.62	0.96
	Peculiarity—Action one doesn't usually engage in	3.00	1.15
4	Milksop—A sympathetic listener, but lacking in understanding	4.19	1.17
	Espionage—Crooked, not truthful	4.12	1.09
	Seclude—To put somewhere in the dark	3.81	1.11
5	Armory—Combined form of some sort of organization	4.94	0.77
	Juggler—Acts in front of a person, respects himself as a juggler	5.00	0.82
	Espionage—A type of sinful devilment	5.44	0.89
6	Nail—Metal I guess, let's say a metal which is made scientifically for purpose of good and bad use	5.75	0.93
	Armory—Part of army subject to call without banner	5.94	0.85
	Diamond—A piece of glass made from roses	6.44	0.63
7	Cushion—To sleep on a pillow of God's sheep	6.75	0.45
	Fable—Trade good sheep to hide in the beginning	6.81	0.40
	Guillotine—Part of law subject only to those without call to stay on earth	6.62	0.62

Table 2

Comprehension Scale

Scale Point	Response	Mean	*SD*
1	Envelope—Deposit it in the mail box	1.06	0.25
	Taxes—Taxes are necessary to support the government	1.06	0.25
2	Land in the city—Because they got more accommodations in the city than in the country	2.00	0.96
	Envelope—Best thing is to bring it to post office	2.00	0.73
3	Envelope—Pass it by or mail it	2.87	0.81
	Theater—Turn in an alarm so that everyone wouldn't get burned up	3.00	0.82
4	Marriage—Proof and identification so you wouldn't get someone else's wife	4.50	1.03
	Shoes—Probably just tradition, Dutch use wood	3.69	1.25
5	Laws—It is reasonable for a group of people to come to some agreement and acceptance of a common good and to aid what has proven to be the best for the many; that is they are made to prevent illegal activities	4.87	0.96
	Shoes—Because leather has undoubtedly proved to be the most durable of all that which has been utilized for the preservation of the feet and to continue the comfort of those, that is the people who have chosen to wear shoes	5.06	1.12
6	Marriage—For ownership you might say and to take care of each other according to health	5.87	0.96
	Marriage—Some people get married in church and some people get married outside of church	5.69	1.09
7	Forest—I'm not good at telling directions. Just walk uphill and when you get to the top it is easier going down	6.31	0.71
	Marriage—For scientific purposes and for the identification of siblings, siblings of the association of the parents	6.31	1.01

to justify construction of rough scales from the obtained data.

The items for the scales were selected by choosing those responses whose means were closest to the desired scale point and whose standard deviations were the lowest available. Differences between the mean rating for each item on the test-retest were tested using Student's t for correlated data (3, p. 277), and for variances using the formula given by McHugh (5) for comparing two correlated sample variances. The few items which showed significant differences were discarded. Three scales for Vocabulary and two for Comprehension were selected, based on the above-mentioned criteria for item selection. The ratings on the first testing are used as it is felt they are more representative of the "true" value, ruling out the effects of practice or memory, negligible though they were. These are presented in Table 1 where all three Vocabulary items for each scale point are listed together, as are the two items at each scale point for Comprehension in Table 2.

Discussion

It will be noticed that in the case of Comprehension there is some indication of a shortening of the scale at the upper extreme. Apparently, our severe Comprehension items were regarded as somewhat less extreme than were the Vocabulary items. The reasons for this are unknown. Our figures on the reliability of clinicians' judgments with this type of material are higher than those we reported previously (4). This may be attributable to sampling, but we feel that the modification and clarification of our instructions may be a prime factor influencing this change. As one of us has said previously, "When dealing with experts in a judgmental situation, the task should be well defined and the criteria set forth clearly. Otherwise the riches of knowledge may yield confusion rather than clarity" (1, p. 274).

The scales here presented are admittedly rough. No attention has been paid to such theoretical problems as equality of interval, etc., and a cross validation of the data is necessary. In their present stage, however, we feel they are valuable as pedagogical aids. Future development of such scales might make it possible to introduce quantitative scoring for the qualitative cues that clinicians presently derive from test materials. This type of material might also be used as a "test" of clinical ability if the materials were given to clinicians, and they were asked to rate them. Individual agreement with the group consensus then might be a measure of clinical aptitude, or at least of an approach to accepted current clinical interpretation.

Received October 19, 1954.

References

1. Arnhoff, F. N. Some factors influencing the unreliability of clinical judgments. *J. clin. Psychol.*, 1954, **10**, 272–275.
2. Asch, P. The reliability of psychiatric diagnoses. *J. abnorm. soc. Psychol.*, 1949, **44**, 272–276.
3. Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
4. Hunt, W. A., Arnhoff, F. N., & Cotton, J. W. Reliability, chance, and fantasy in inter-judge agreement among clinicians. *J. clin. Psychol.*, 1954, **10**, 294–296.
5. McHugh, R. B. The comparison of two correlated sample variances. *Amer. J. Psychol.*, 1953, **66**, 314–315.
6. Mehlman, G. The reliability of psychiatric diagnoses. *J. abnorm. soc. Psychol.*, 1952, **47**, 577–578.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Differential Prediction of a Specific Behavior from Three Projective Techniques

Claire M. Vernier,
VA Center, Martinsburg, West Virginia

J. Frank Whiting,
VA Hospital, Rutland Heights, Massachusetts [1]

and Malcolm L. Meltzer
Catholic University of America

In recent years, there has been an increasing interest on the part of psychologists in the use of projective techniques for prediction of behavior which has important social consequences. Such study not only has the value of contributing crucial data necessary for an integrated theory of behavior in general, and projective techniques in particular, but also enables psychology to fulfill its professional function of aiding the culture in maintaining and helping its emotionally disturbed members.

The area of behavior with important social consequences which was chosen for the current study was the great number of tuberculous patients who leave hospitals and sanatoriums against medical advice (subsequently referred to as AMA behavior). Since the overwhelming majority of these patients have active, communicable disease, the social consequences of their leaving the hospital are infinitely more serious than would be the results of similar actions on the part of neurotics or persons suffering from noninfectious physical diseases. These latter individuals present difficulties to themselves alone or at the most to their immediate family and friends. The actively tuberculous patient is a potential source of infection to the entire community as well as an obvious danger to himself in having interrupted or terminated his treatment.

A review of the literature on tuberculosis reveals several important facts about this problem of AMA behavior. First, it is a world-wide problem and not merely limited to certain hospitals with inadequate facilities. Second, the consensus in the literature is that such action stems from pressures from three sources: (*a*) the person's own needs, (*b*) the influence of friends and relatives outside the hospital, and (*c*) pressures put upon the person by the hospital environment. Third, while there has been much speculation concerning personality in the tuberculous, including efforts to define a tuberculous personality or character type, there has been little in the way of controlled experimental investigation about any phase of the psychological concomitants of tuberculosis. Only Gurel and Jennings (**5**), to our knowledge, have attempted any integration of theory with empirical investigation of this problem. Essentially the theoretical basis of their study was one of a temporal relationship between the operation of reward and punishment and the acquisition and retention of behavior (**9**). Thus, in terms of tuberculous patients, those who exhibit AMA behavior demonstrate an inability to inhibit impulsive behavior in order to derive long-term satisfaction, while patients who remained in the hospital and were discharged with maximum hospital benefit (subsequently referred to as MHB behavior) were able to do so. More general consideration of the implications of this theory is beyond the scope of the present paper. However, the above hypothesis concerning the process determining patients' behavior is the general hypothesis, at the level of behavior theory, on which the present study is based.

The particular projective techniques utilized in this study were the Rorschach test, the Thematic Apperception Test, and House-Tree-Person Test. The particular hypotheses tested experimentally will be described in the following sections. It suffices here to point out that these three tests constitute samples of projective techniques which vary both in terms of instructions and degree of structure of overt stimulus material presented to the subject.

---

[1] Formerly Clinical Psychologist, VA Center, Martinsburg, West Virginia.

Subjects and Method

Since January of 1953 a group psychological testing program has been carried out for all admissions to the TB Service as soon as they have been cleared by the ward physicians as physically able to take the tests. The major purpose of the testing program is to provide significant information to the physicians in regard to the patients' attitudes toward their illness, hospitalization, and treatment, and to identify emotional problems which could interfere with hospital adjustment or response to treatment. However, the test protocols are also available for research purposes. As of January, 1954, approximately 350 patients had been tested. Of this group, 63 had received MHB discharges and 95 had received AMA discharges.

From these two samples of discharged patients, all cases were chosen who had been given any one of the three tests selected for analysis. Inasmuch as the particular battery of psychological tests used for the admission testing program varied according to the preferences of the staff psychologist assigned to the TB Service, the populations for study of the three tests overlap but are not identical. Description of the specific subjects used for each test is included in later sections.

Analysis of the Tests

Rorschach

The hypotheses tested in this study by means of the Rorschach were derived from a series of deductions based on the writers' general hypothesis at the level of behavior, their assumptions concerning the projective technique situation, and the generally accepted clinical interpretations of scoring categories of the Rorschach responses.

Since the basic theory being tested was that AMA patients are significantly less able to inhibit impulsive behavior in order to obtain long-term satisfaction than are MHB patients, only the determinant categories of the Rorschach were chosen for analysis. According to Rorschach literature (6), the determinants reflect the emotional aspects of personality and serve as a principal basis for predicting overt behavior.

The following hypotheses in regard to the determinants were formulated:

a. AMA patients will be higher than MHB patients in: $F-$, $F-\%$, CF, C, ΣC, FM.

b. MHB patients will be higher than AMA patients in: F, $F\%$, FC, M, C', c, k, K.

c. Within the AMA group, $CF+C$ should exceed FC, $FM+m$ should exceed M, and $FM+m$ should exceed $c+C'$.

d. Within the MHB group FC should exceed $CF+C$, M should exceed $FM+m$, and $c+C'$ should exceed $FM+m$.

Subjects. Rorschachs were selected for 46 male AMA patients and 46 male MHB patients so that the two groups were comparable for age, race, marital status, education, degree of disease, and type of treatment prescribed. Respective mean ages were 39.65 and 39.67. All tests for these 92 patients, identified only by a number to avoid possible bias, were rescored by one of the authors to maintain consistency.

Data analysis. The first score to be examined was the mean number of responses for each group (R). Had there occurred a significant difference between the number of responses for the two groups, raw determinant scores could not have been used and the effect of R on each of the several determinant categories would have had to be controlled statistically. However, mean R for the AMA and MHB groups was 14.5 and 14.3, respectively, and the chances were 99 out of 100 that such a difference would arise by chance alone.

Two of the determinants, C and K, did not occur sufficiently often for the purposes of testing the empirical hypotheses by means of an analysis of differences between the AMA and MHB groups. Such results are in line with previous Rorschach experience.

The remainder of the determinants analyzed produced sufficient data for the purposes of testing the hypotheses formulated. Since the empirical hypotheses were stated in terms of a specific direction of the results, a one-tailed *t* test was used to determine probabilities (4). The results are listed in Table 1 and Table 2.

Table 1

Comparison of AMA and MHB Groups on Rorschach Determinants

($N = 92$)

Determinant	Mean		*SD*			
	AMA	MHB	AMA	MHB	CR	*p*
AMA predicted greater						
$F-$	1.9	1.8	2.4	2.4	0.2	.45
$F-\%$	33.0	25.0	25.0	18.0	1.8	.04*
CF	1.8	1.4	2.4	2.1	1.0	.16
ΣC	2.5	2.1	3.0	3.0	0.7	.25
FM	1.5	2.0	2.1	2.8	1.0	.16
m	0.7	0.7	1.1	1.3	—	—
MHB predicted greater						
F	5.0	4.9	6.4	5.9	0.1	.47
$F\%$	45.0	48.0	23.0	22.0	0.6	.27
FC	0.6	1.0	1.1	1.5	1.3	.10*
M	0.9	1.3	1.3	1.7	1.3	.10*
C'	0.7	1.2	1.2	1.8	1.7	.04*
c	1.3	1.5	2.0	2.0	0.5	.31
k	0.7	0.5	1.2	1.0	1.0	.16

* Such differences could occur by chance 10 or less times in 100 instances.

In making a series of tests of empirical hypotheses one can always expect that a certain number of significant differences can arise due to chance relationships in sampling. Therefore, a chi-square test was made between the observed number of significant results and near significant results ($p < .10$) and the expected number of such results. Seventeen tests were performed and thus 1.7 number of statistically reliable differences would be expected by chance alone. Eight significant results were actually obtained. The chi-square test of the difference between the observed and expected frequencies yields a value of 51.2. With 1 degree of freedom the chi-square value must be greater than 6.635 in order to be beyond the .01 level of probability. Therefore, it is concluded that the observed number of statistically reliable differences could not have developed from chance alone.

Conclusions. One-half of the predicted Rorschach scores and ratios, assumed to be indicative of ability to control behavior, were found to differentiate between the two groups at better than chance level. These findings would support the general hypotheses postulated; however, none of the obtained differences for the individual Rorschach determinant categories are significant at a level which would yield accurate prediction for the individual case.

Thematic Apperception Test

The second approach to the problem of the prediction of AMA behavior of tuberculous patients was the utilization of a projective technique designed primarily to elicit underlying personality dynamics. It was hypothesized that content analysis of the Thematic Apperception Test would reveal certain needs operating which were common to the AMA group but absent in MHB patients. Should this be true, an explanation of AMA behavior could be made, which in turn would allow for prediction and for institution of preventive measures.

Consideration of the nature of the disease and of treatment methods suggests that two important problem areas confronting the TB patient are those of aggression and dependence. The patients who are able to remain hospitalized for lengthy periods of time and undergo painful treatment may display strong needs for punishment and for submission and dependence. The AMA patients, on the other hand, seem to show an inability to accept either the loss of autonomy or the enforced dependence necessary for treatment, and by their antisocial behavior seem to reflect needs

Table 2

Comparison of AMA and MHB Groups on Rorschach Ratios

(Predicted AMA group should exceed MHB group on each.)

Ratio	Percentage of subjects		CR	*p*
	AMA	MHB		
$FM+m>c+C'$	70	57	1.3	.10*
$FM+m>M$	67	52	1.5	.07*
$\Sigma C>M$	70	50	2.0	.03*
$CF+C>FC$	57	39	1.8	.04*

* Such differences could occur by chance 10 or less times in 100 instances.

to direct aggression outwardly. It was therefore hypothesized that the AMA group would score high on needs for externalized aggression, dominance, and independence. The MHB group was expected to display the converse.

One further analysis of the TAT records appeared appropriate for prediction of AMA behavior. It was hypothesized that patients who leave AMA also are more prone to act on their needs than are those patients who remain for the full course of treatment. Action was scored from the TAT stories in terms of whether the need expressed took the form of direct action or whether the need was expressed only as fantasy, contemplation, or as a possibility in the future. The first instance, where the need found expression in action, was characterized as the motor level; when the particular behavior was only contemplated or fantasied it was designated as the premotor level.

Subjects. Twenty white male veterans who had been given TAT's as part of the admission group testing program were selected. From their history 10 were identified as AMA and 10 as MHB. None of the AMA patients had been able to remain in the hospital for more than 4 months, the mean length of hospitalization being 2.73 months. All of the MHB patients, on the other hand, had been in the hospital for at least a year, the mean length of hospitalization as of June, 1954, being 16.4 months. The mean age of the AMA group was 35.1 yrs.; that of the MHB's, 33.4 yrs. The two groups were also equated as to economic level, pension status, marital status, stage of disease, and type of treatment.

Data analysis. Four cards of the TAT had been regularly given to each of the twenty subjects. These were numbers 1, 4, 6BM, and 7BM. The TAT records were scored by the need-press system devised by Aron (2). A ratio was computed for each patient by totaling the number of needs for external aggression, dominance, and independence (EDI) and relating this to the total number of needs for inward aggression, submission, and dependence (ISD). The data were placed in a fourfold table and analyzed by the chi-square technique.

Table 3

Comparison of AMA and MHB Groups on the TAT ($N = 20$)

Condition	χ^2	p
EDI vs. ISD with AMA and MHB combined	0.80	.30
AMA vs. MHB on EDI and ISD	0.83	.30
Motor vs. Premotor with AMA and MHB combined	0.20	.50
AMA vs. MHB on Motor and Premotor Responses	0.21	.50

Each need was scored also on the motor-premotor dichotomy. These data were placed in a fourfold table and subjected to a chi-square analysis.

It was determined first that the occurrence of EDI responses does not differ significantly from the number of ISD responses with the samples combined. Should one of the variables predominate in all the TB subjects, it would be more difficult to discover a significant difference between the subgroups. However, as can be seen in the first result of Table 3, neither EDI nor ISD responses are significantly more plentiful when the two groups are combined. Chi square is 0.80 and this fails to reach even the .30 level of significance.

The second result in Table 3 relates more directly to the hypothesis that AMA patients display strong needs for external aggression, dominance, and independence, while MHB patients have greater needs for internal aggression, submission, and dependence. After characterizing each subject as EDI or ISD and comparing the number of such individuals in each group, we find that chi square is only 0.83. This does not reach the .30 level of significance and therefore fails to confirm our major hypothesis concerning the dynamics operating in tuberculous patients who leave the hospital against medical advice.

Our last analysis concerned the level of response, i.e., whether the need was expressed as definitely carried out in motor behavior, or whether it was presented in the form of a wish, fantasy, or intention and therefore was premotor. The third result in Table 3 shows that the number of needs scored as at the motor level did not differ significantly from the number of needs scored at the premotor level with the AMA and MHB samples combined. Chi square is .20 which is not significant at the .50 level of probability. Once again, if significant differences do exist between the groups, they should not be obscured.

However, the AMA and MHB groups are not differentiated by the level of response as can be seen in the final result of Table 3. Chi square is only .20, and, therefore, there are 50 chances out of a hundred that the difference is a chance one alone. From this sample then, a preponderance of responses at the motor level does not of itself indicate a propensity for AMA behavior, since responses at this level are fairly evenly divided between both groups.

Conclusions. The failure of the TAT to effectively discriminate between the AMA and MHB patients can be viewed from several aspects. Specifically, it is quite possible that our particular hypotheses concerning the dynamics of tuberculous patients are in error. Furthermore, it can be questioned whether any definite constellations of needs or dynamics are operating in tuberculous patients in general or AMA patients in particular. It has long been an axiom of the psychodynamic approach that the same bit of behavior often reflects different needs in different individuals or even distinct needs in the same person at separate times. Should this be so in relation to the AMA problem, the use of tests revealing dynamic needs and content of wishes and strivings becomes suspect as an adequate tool for the prediction of this particular bit of behavior.

One exception to the last statement might be pointed out. As yet little is known about content-type tests which present the subject with a stimulus which symbolizes the particular situation in which the prediction of overt behavior is to be made. What is being considered by the authors is a modification of the TAT which would depict the particular problem areas and situations facing the TB patient. Perhaps such a focusing could elicit needs and unconscious attitudes more germane to the overt behavior to be predicted and some communality could be discovered. This particular problem awaits further investigation.

House-Tree-Person Test

Inclusion of a drawing test in the battery of techniques selected for analysis was based on the hypothesis that nonverbal tests tend to yield data less subject to consciously controlled evasion and censorship and thus provide a more efficient basis for prediction of impulsive, overt behavior (11). The HTP test was chosen since the house and the person constitute stimuli approximating two of the focal elements in the problem behavior being studied, namely institutional versus home living, and the patient's perception of his own goals and needs.

Contrary to the usual assumptions concerning the house drawing as reflecting attitudes toward home and family, the authors assumed that the house drawn by the patient would symbolize aspects of his concept of and attitudes toward his important living space (3). The following specific hypothesis was made: AMA and MHB patients will differ significantly in the treatment of all parts of the House drawing which pertain to an interaction of outside and inside. The following 13 test characteristics were thus selected for differential analysis:

1. door placement
2. window details
3. details on door
4. number of windows
5. chimney treatment
6. inclusion of walk
7. inclusion of steps
8. inclusion of fence
9. inclusion of road
10. inclusion of garage
11. inclusion of shrubbery and trees
12. direction of house facing (e.g. differential placement of small side)
13. ground line drawn

Three aspects of the person drawing were chosen for analysis: direction the person was facing, type of neck treatment, and presence or absence of hands. Underlying the use of these three variables were the current theories of drawing interpretation: (*a*) facing to the right hand side of the paper represents externalization of action; (*b*) the neck depicts the type of controls over body impulses; (*c*) treatment of hands depicts aspects of interpersonal relationships (**8**, **11**).

Subjects. Thirty male MHB patients and 50 male AMA patients were selected. Mean ages for the two groups were 36.2 and 34.7, respectively, with ranges of 18–65 and 21–63.

Approximately 85% of each group were white (83% and 86%, respectively); the remainder were Negro. Total length of hospitalization for the MHB group ranged from 2 to 47 months, with a mean of 11.0 months; for the AMA group the mean was 3.8 months, with a range of 1 to 9 months. The two groups were equated for education, occupational status, marital status, degree of disease, and type of treatment.

Data analysis. Incidence of occurrence was tabulated for each of the 16 test variables selected and the differences in percentages computed. Results are given in Table 4.

Of the 16 variables tabulated, two did not occur sufficiently frequently in either group to permit further analysis. Of the remaining 14, seven yielded differences between the two that could be expected to occur by chance less often than five times in 100. The expected frequency of results below .05 by chance alone would be 0.8. A chi-square test yields a value of 40.61, which is well beyond the .01 level of 6.63 for one degree of freedom.

Table 4

Comparison of AMA and MHB Group on the HTP ($N = 80$)

	Percentage of subjects			
Item	AMA	MHB	CR	p
House				
Door on left side of page only	6	27	2.3	.02*
Door details, 2 or more present	24	30	0.6	.55
Window detail drawn as +	62	23	3.9	.0001*
No. of windows, 4 or more	32	53	1.8	.08
Smoke from chimney	26	0	4.3	.0001*
Walk present	20	10	1.2	.21
Steps drawn	18	27	1.0	.32
Road indicated	12	0	2.4	.02*
Shrubbery and/or trees drawn	2	10	2.0	.06
Small side of house on right side of page	44	20	2.4	.02*
Ground line drawn	22	3	2.7	.01*
Person				
Head facing to right of page	20	7	1.6	.11
Neck either omitted or drawn wider than long	50	37	1.2	.23
Hands omitted	44	23	2.1	.02*

* Such differences could be expected to occur by chance less often than three times in 100.

Conclusions. The general hypothesis was confirmed that a nonverbal test which presented stimuli to the subjects, certain elements of which were directly related to the problem area of behavior to be predicted, would yield differences between the two groups sufficiently reliable to permit prediction for the individual case.

Application of Results to a Second Sample of Patients

Patients admitted for the first time to the hospital and seen for psychological testing during the period of January through March, 1954, constitute a group available for a partial approach to cross validation of the findings. Test reports written by the author assigned to the TB Service at that time included a specific prediction of whether the patient was a potential AMA risk. Since the report also included a description of the factors involved for the individual and the implications for treatment, it is impossible to evaluate the influence of the report, the prediction, and the suggestions upon the subsequent actions of the patient. The psychologist responsible for the predictions had knowledge of the findings from the statistical analysis of the three tests but did not use any type of index or quantified approach. The analysis of the accuracy of the predictions is shown in Table 5.

Since all the patients had been in the hospital less than six months, these data are not final figures. It is interesting to note that to date all of the errors in prediction have been in one direction, namely, patients identified as potential AMA risks who still remain in the hospital.

General Discussion of Results

Implications of findings for the problem of AMA behavior prediction. Despite the small samples available for analysis, a sufficient number of reliable differences in test scores were found to permit accurate prediction for individual patients. The data for one of the

Table 5

Application of HTP Findings to First Admission Patients Over a Three-Month Period ($N = 26$)

Prediction	Number of patients	Number of patients discharged AMA*
An AMA potential	16	12†
No AMA potential	10	0†

* As of August 15, 1954.

† Combined accuracy of prediction equals 85 per cent.

tests, the drawing of a house, justified the development of a specific index. Use of this predictive index with two other groups of TB patients will be reported subsequently.

While the findings support the general hypothesis that the patients who leave the hospital AMA show less control over their behavior than do patients who stay, there was no confirmation of specific differential dynamics or personality characteristics between the two groups. The one concrete conclusion which may be drawn is that, regardless of the problems and motivations of the individual patient, those cases who leave against advice project their perception of ease of access and exit into their representation of a dwelling.

Since the study demonstrates that potential AMA patients may be identified accurately shortly after their first admission to the hospital, the next problem is that of preventive treatment. The major implication of the present findings is that one important factor is the perception of the hospital as an empty and unsatisfying place which has no barriers to prevent departure. Thus, one area of needed attitude change to prevent AMA discharge would appear to be the patients' conceptions of the hospital itself.

Implications of results for projective test theory. For several decades psychological literature has presented discussion of various aspects of the problem of prediction of behavior from projective techniques. Much of the attention has been focused on difficulties in validation of the tests or verification of isolated hypotheses about selected test scores or characteristics. Within recent years increasing emphasis has been placed on the use of projective measures in the practical area of evaluation of changes occurring during the psychotherapy process and the prediction of response to psychotherapy (1, 10).

However, the absence of an integrated, comprehensive theoretical basis continues to handicap much of the clinical research with projective methods. Although there has been an increasing tendency to rely on field theory as an adequate frame of reference for the ordering of projective behavioral data, relatively minimal use has been made of this conceptual framework in experimental studies. The authors have derived their basic hypothesis of the use of interaction measures for behavior prediction from the basic postulate of field theory that behavior is a function of the person-situation relationship (7). Therefore, to predict reliably a specific behavior, the psychological instrument used must measure the interaction of the person's needs and the situational stimuli. To the extent that a given projective test provides such measures, the more accurate the prediction of the externalized behavior.

In selecting the three diverse types of projective techniques for analysis in this study, the authors had assumed that a differential level of accuracy in prediction would result. While all three of the tests provide measures of individual dynamics, the tests may be arranged on a continuum of the amount of situational stimuli presented. In terms of the very literal description of the behavior to be predicted, namely, a person walking out of a living place, the tests could be ordered as follows: drawing of a tree, TAT, Rorschach, drawing of a person, and drawing of a house.

The findings of the present study support this hypothesis. Neither the tree drawings nor the TAT yielded any reliable differences between the two groups of patients. The Rorschach and person drawings gave reliable group differences but could not be used for clinical prediction. The house drawings were successful in prediction for the individual patient.

The major implication of the present study for projective methods would appear to be a confirmation of the importance of analysis of interaction between person and situation for accurate behavioral prediction. For such an analysis to be possible, it is essential that the tests, while ambiguous, present stimuli which tap the specific situation or area in which the behavior to be predicted occurs.

Summary

Three projective tests were analyzed in terms of reliability of differentiation between two groups of hospitalized TB patients. The two groups of patients were comparable for age, race, education, marital status, pension status, degree of disease, and type of treatment; they differed only in length of stay in the hospital and the type of medical discharge received.

The TAT yielded no significant differences between the two groups and none of the specific hypotheses were verified.

Eight of the determinant scores and ratios on the Rorschach test differentiated reliably between the two groups, but not at a level which would permit clinical prediction. While the patients differ significantly in degree of control over personal impulses, this factor does not in itself provide a sufficient explanation for the behavior under study.

Drawings of the house provided seven variables which differentiated the two groups; two of these differences were at a level which would permit prediction for an individual patient. Subsequent application of these findings to a second sample of TB patients resulted in accurate prediction for 85 per cent of the cases.

The differential accuracy in prediction for the three projective tests studied would support the authors' basic hypothesis that accurate prediction of specific overt behavior from projective techniques is dependent upon the extent to which the test provides a measure of the interaction between needs of the individual and a symbolization of the external factors of the situation in which the behavior occurs.

Received September 22, 1954.

References

1. Abt, L. E., & Bellak, L. *Projective psychology.* New York: Knopf, 1950.
2. Aron, Betty. *A manual for analysis of the Thematic Apperception Test.* Berkeley, Calif.: Willis E. Berg, 1949.
3. Buck, J. N. The H-T-P technique; a qualitative and quantitative scoring manual. *J. clin. Psychol.,* 1948, **4**, 317–396; 1949, **5**, 37–76.
4. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Gurel, L., & Jennings, T. S. The use of psychological tests to predict irregular discharge in the hospitalization of the tuberculous. Paper read at East. Psychol. Ass., Boston, April, 1953.
6. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
7. Lewin, K. *Field theory in social science.* New York: Harper, 1951.
8. Machover, Karen. *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1949.
9. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.
10. Vernier, Claire M. Current avenues of psychological research in projective techniques. *Quart. Rev. Psychiat. Neurol.,* 1952, **7**, 1–4.
11. Vernier, Claire M. *Projective test productions.* Vol. 1. *Projective drawings.* New York: Grune & Stratton, 1952.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

The Psychiatric Validity of an Inverted Factor Analysis of Rorschach Scoring Categories

A. W. Bendig
University of Pittsburgh

and Roy M. Hamlin
University of Pittsburgh, Western Psychiatric Institute and Clinic

Numerous studies have attempted to apply statistical techniques, such as multiple regression and the discriminant function, to the problem of the psychiatric validity of the Rorschach scoring categories. Typically these studies attempt to isolate a subgroup of categories or elements from the total profile and to estimate a series of weights to be applied to these elements which, in linear combination, will best discriminate between psychiatric classificatory syndromes. This purely psychometric approach, which minimizes the role of the clinical interpreter of the Rorschach protocol, can be successful when one necessary condition is met, i.e., when *S*s alike in a selected group of clinical behaviors (psychiatric syndrome) are alike on a selected group of Rorschach profile elements and when these *S*s differ on the same elements from *S*s in some other psychiatric category. Failure to find diagnostic validity of the Rorschach using this approach may be attributable to several possible factors. The psychiatric classification method may fail to provide internally homogeneous and externally heterogeneous categories of behavior, and result in considerable behavioral overlap between the psychiatrically distinct categories. Or several distinct clusters of Rorschach elements may occur within a given diagnostic category, but these clusters are still different from those within another psychiatric category. Or the element clustering may cut across diagnostic lines with *S*s from different categories being more alike Rorschachwise than are *S*s within the same categories. These last two possibilities are suggested by the study of Wittenborn and Weiss (5) who reported an inverted factor analysis of the intercorrelations on 55 symptom-rating scales of 20 *S*s with the same psychiatric diagnosis (manic-depressive psychosis—manic state). The absence of a general factor from the six extracted leads these authors to conclude that important subgroup differences exist within such a homogeneous psychiatric group and that these differences are obscured by the similar diagnostic label. Finally, the Rorschach elements may be interindependent and no clustering may occur.

One promising preliminary approach to the isolation of clusters of Rorschach elements is the inverted factor analytic or *Q* technique. Factors are extracted from a matrix of intercorrelations between *S*s (rather than from a matrix of intercorrelations between variables as in the standard *R* technique) and, after the factors are rotated to simple structure, the factor loadings used to group the *S*s. Cattell (2, pp. 502–503) regards this method as especially useful in isolating subgroups within a nonhomogeneous population and says the simple structure factors should correspond to these population variables when the dependent variable measures are affected by the population variables (2, p. 511). This suggests that if we selected groups of *S*s from distinctly different psychiatric classifications, selecting *S*s for internal behavioral homogeneity and intergroup heterogeneity, and factor analyzed the correlations between *S*s over a large series of Rorschach elements, the factor loadings would permit a psychometric grouping of the *S*s. If this grouping corresponded to the psychiatric grouping we would have evidence that a psychometric technique could profitably be applied to a larger sample of *S*s.

Procedure

Subjects. Sixteen *S*s were used with four *S*s in each of four diagnostic categories. Group 1 *S*s had all been diagnosed as paranoid schizophrenics with a minimum of disagreement in regard to diagnosis. In each case delusions had been clearly elicited and there was a history of markedly unusual social behavior. Clinical material gave evidence of unusual difficulty in sexual adjustment and confusion in sexual role or homoerotic tendencies. Group 2 *S*s had all been diagnosed as imbeciles. There was no evidence in any case of brain injury or of special physiological determinants. Group 3 had a diagnosis of anxiety state. Clinical material gave evidence of marked current anxiety at the time of testing. In no case was there any evidence of schizophrenic personality features, bizarre thinking, or loss of affect. History material demonstrated strong discomfort and incapacitation from anxiety feelings. All four *S*s were psychiatric patients. Group 4 *S*s all had a diagnosis of involutional melancholia. Clinical notes (dated shortly before and after testing) indicated that each *S* was clearly depressed at the time. Feelings of inadequacy, futility, and guilt were verbalized. The gross change in affect was the primary feature of the illness.

Several hundred case records were reviewed in selecting *S*s. The purpose of the selection was to insure that *S*s in any one of the four groups not only had a similar diagnosis, but were also as homogeneous as possible in terms of the clinical picture and current status. Controls reflected this purpose. For example, the four *S*s in Group 1 were all male; the four in Group 4 all female.

Rorschach. Each protocol was scored three times, once by the original examiner and independently by two other scorers for purposes of this study. Totals for variables used in the present study were compared for the two independent scorings and inspection indicated no disagreements that would affect the later dichotomization of the categories. The selection of *S*s and scoring was under the direction of the second author who turned the data over to the first author for the statistical analysis with all identifying information (such as diagnosis, sex, etc.) removed. The first author was not aware of the specific diagnostic groups involved nor the groupings of the *S*s until after the factor analysis and subsequent rotation of factor axes had been accomplished.

Results

The distribution of the 16 patients on each of the 42 Rorschach elements was found and each element dichotomized as close to a median value as was possible. Patients scoring above the median on an element were scored "plus": those falling below the median score for that element were scored "minus." Thus for each patient we had a series of 42 plus and minus ratings. Each patient was correlated with every other patient by comparing their plus and minus ratings on each of the 42 elements and fourfold tables constructed reflecting the correspondence between their ratings. Phi correlation coefficients were computed from each of the 120 fourfold tables since the assumption of normality necessary for the use of tetrachoric correlations was obviously not possible for the distributions of many of the Rorschach elements.[1]

The matrix of intercorrelations between the 16 *S*s was factor analyzed by the standard centroid method, using the largest correlation in each column as the communality estimate at each stage of the extraction process. The process was terminated after the third factor had been estimated because of the small remaining residual correlations (the median of the absolute magnitudes of the residuals was .07). The original three factors were graphically and analytically rotated to achieve, as

---

[1] The list of Rorschach elements, ranges of responses scored plus, and numbers of patients falling above the dichotomization points, and the phi intercorrelations between the patients and residual correlations after extraction of three factors have been deposited with the American Documentation Institute. Order Document No. 4503 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Table 1

Factor Loadings (Decimal Points Omitted) of Three Factors Extracted from Phi Intercorrelations of Rorschach Profiles of Psychiatric Subjects

		Productivity variables		Factor loadings						
				Original			Rotated			
Psychiatric group	Patient number	*P*	*PQ*	I	II	III	A	B	C	h^2
Neurotic (Anxiety)	20	67	22	43	−32	26	54	14	23	35
	40	64	23	52	−54	15	66	−02	39	58
	46	50	25	−37	−28	−34	−12	−56	−06	33
	55	40	24	−38	19	−30	−45	−23	−15	27
Depressed	4	45	25	45	−24	−14	24	−01	47	28
	23	67	22	−13	−45	17	39	−27	−15	25
	31	50	25	−39	14	40	00	12	−57	33
	54	62	24	41	−10	−22	08	02	47	23
Paranoid schizophrenic	3	79	17	12	−09	27	27	15	−06	10
	22	17	14	−18	64	24	−39	47	−36	50
	24	62	24	−35	−34	−07	09	−44	−21	24
	29	36	23	39	16	−28	−15	15	46	26
Imbecile	14	33	22	77	10	−08	14	41	65	61
	37	24	18	20	36	36	01	54	−10	30
	53	14	12	24	51	28	−13	61	−03	40
	63	29	21	−27	−18	47	32	00	−47	33
	Per cent of total variance			15	11	8	10	11	13	34

far as was possible, orthogonal simple structure. Both the original and rotated factors tended to be bipolar, eliminating the possibility of positive manifold, and simple structure could not be completely achieved.

Cattell (1, pp. 321–327) has noted that the statistical characteristics of the intercorrelated variables, such as test difficulty and variance, can affect the magnitudes of the intercorrelations and factor structure. Variables of either extremely high or low difficulty level show small variance, or are high in "eccentricity." This is particularly important when point correlations, such as the phi coefficient, are used. At the empirical level, Wittenborn (4) has shown that many Rorschach element scoring categories are highly related to the total number of responses given by the *S*. In the present study this background suggests that the factor loadings of a given *S* may be a function of the proportion of his elements scored "plus" which in turn is a function of his total number of Rorschach responses. To control this possibility two additional measures related to the Rorschach "productivity" of each *S* were computed: the percentage of Rorschach elements scored "plus" for the *S* (*P*), and the "eccentricity" or variance of this percentage (*PQ* where $Q = 100 - P$).

In Table 1 are given the original and rotated factor loadings for each of the 16 *S*s and also the communalities (h^2) and productivity variables (*P* and *PQ*) for each *S*.[2]

To test the above noted hypothesis of relations between factor loadings and productivity variables, product-moment correlations were computed between each of the original and rotated factors and *P* and *PQ*. The obtained correlations, along with the intercorrelation of *P* and *PQ* and the multiple correlation of *P* and *PQ* with the factor loadings, are given in Table 2. Factors II and B were both negatively correlated with *P* and *PQ*, while factor A was positively related to *P* and factor III

---

[2] The transformation matrix for converting original to rotated factor loadings has been deposited with the American Documentation Institute. See footnote 1.

Table 2

Product-Moment Correlations Between Factor Loadings and Productivity Variables

Variables		*P*	*PQ*	Multiple *R*
Factor	I	.03	−.07	.10
	II	−.79**	−.64**	.83**
	III	−.12	−.52*	.55
	A	.60*	.22	.61*
	B	−.58*	−.71**	.75**
	C	.16	.23	.23
h^2		−.42	.14	.59*
PQ		.52*		

* Significant at the .05 level.
** Significant at the .01 level.

was negatively associated with *PQ*. The multiple correlations of *P* and *PQ* with factors II and B were significant at the .01 level ($R =$ .83 and .75) and with factor A and the communality variable (h^2) at the .05 level ($R =$ .61 and .59). The multiple correlation of *P* and *PQ* with factor III ($R = .55$) was close to statistical significance at the .05 level. Since 10 of the 21 correlations were significant at the .05 level or better, the hypothesis that factor structure was affected by individual differences between *S*s in Rorschach productivity appears confirmed.

The original hypothesis of this study was that one or more of the factors extracted from the intercorrelations of Rorschach profiles of the 16 *S*s would correspond to their psychiatric classification. To test the validity of the loadings in discriminating among the psychiatric groups, curvilinear correlations (eta) were computed between the diagnostic grouping of the *S*s and nine variables; the six original and rotated factors, the communality (h^2), and the two productivity variables (*P* and *PQ*). In addition, the loadings on the six factors and the communality were corrected for the influence of *P* and *PQ* (see Table 2) by computing a multiple regression equation for predicting loadings from the obtained *P* and *PQ* values for each *S* and subtracting these predicted loadings from the obtained loadings. Validity coefficients (eta) were then calculated for these corrected values. The uncorrected and corrected eta coefficients can be found in Table 3. Both of the productivity variables, *P* and *PQ*, discriminated among the psychiatric groups only at the .10 level (eta = .66 and .64). Comparison of the means of the four groups indicated that the imbecile group was low on both *P* and *PQ* with the other three groups showing similar means. Rotated factor B also discriminated among the groups at the .10 level (eta = .63). However, the comparison of the group means showed the imbecile group to be high on this factor (the other groups again showing homogeneous means) and the multiple correlation of factor B and the productivity variables (*P* and *PQ*) to be quite large ($R = .75$). When factor B loadings were corrected for their relationship to *P* and *PQ*, the validity coefficient fell to an insignificant value (eta = .35). The promising validity of uncorrected factor B was apparently attributable solely to its relation to *P* and *PQ*. Six of the seven uncorrected factor validity coefficients decreased in magnitude when they were corrected for the relation of the factor variables to the productivity variables, while the eta for one factor variable (rotated factor A) increased after correction (from .29 to a still insignificant .54).

Discussion

In general, the results of this application of inverted factor analytic methods to the Rorschach profiles of psychiatric groups are negative. The intercorrelations between the *S*s on the 42 Rorschach elements were relatively

Table 3

Validity Coefficients (Eta) of Factor Loadings in Discriminating Psychiatric Groups

Variables		Uncorrected	Corrected for *P* and *PQ*
Factor	I	.24	.26
	II	.54	.44
	III	.43	.42
	A	.29	.54
	B	.63*	.35
	C	.15	.12
h^2		.48	.35
P		.66*	
PQ		.64*	

* Significant at the .10 level.

low and the three factors extracted accounted for only 34 per cent of the total variability. This is in marked contrast with the results of Wittenborn and Weiss (5) where the extraction of six factors from the intercorrelation of 20 *S*s on 55 rating scale variables accounted for 68 per cent of the total variance. This low factor communality militates against a clear factor structure and consequently against an unequivocal grouping of the *S*s. Many influences may have lowered the communality: low Rorschach element reliability, the type of correlation coefficient used, etc.

The obtained factor structure showed no evidence of validity against the psychiatric category criterion. Only the productivity variables, based on the number of Rorschach responses, demonstrated any relation to the psychiatric grouping and this only because of the inclusion of an imbecile group of *S*s. As found by Wittenborn (4), the obtained factor structure was related to *S*s' productivity and when this relation was statistically eliminated the factor loadings showed no significant relation to psychiatric classification. These results are in marked contrast to those of Chambers (3) who found clinical interpreters of Rorschach protocols to demonstrate significant psychiatric validity. Five clinical *S*s were selected from five behaviorally defined psychiatric categories (the four used in the present study plus a paretic group) with the selection criteria emphasizing unequivocalness of diagnosis and intragroup homogeneity in clinical behavior. Five total Rorschachs, one from each psychiatric group, were submitted to 20 experienced Rorschach interpreters who were requested to psychiatrically categorize these "blind" Rorschach protocols. The highly significant number of correct categorizations gave evidence of the adequacy of clinical judgment in psychiatric diagnosis.

Probably the most adequate hypothesis to explain the differences in psychiatric validity of the Rorschach profile and judgment approaches lies in the additional clinical information contained in a complete protocol that is not present in a profile of traditional Rorschach scoring categories. Basically this added information is drawn from two aspects of the protocol: the specification that a given response was elicited by a particular area on a particular card, and the intracard and intercard sequencing of the responses. These sources of information are ignored in the usual scoring elements, but are commonly noted by experienced clinicians as being highly important in forming their judgments. One suggested advance in the objectification of Rorschach validity would consist of the development of methods of defining and quantifying these card-area, response-sequence, and profile-element interrelationships as additional predictors of psychiatric diagnosis. Another area of investigation lies in curvilinear and discontinuous relations among quantified aspects of Rorschach performance and between these aspects and psychiatric diagnosis. Most of the heretofore developed psychometric methods assume rectilinearity and continuity: assumptions that appear quite hazardous in defining the structure of human personality. Which of these several approaches is potentially most fruitful could be decided by additional studies of how the practicing and experienced Rorschach interpreter makes his judgments and what information is most necessary to him for valid diagnosis.

Summary

Four clinical *S*s from each of four psychiatric diagnostic categories (anxiety neurotic, depressive, paranoid schizophrenic, and imbecile) were selected on the criteria of intragroup similarity in clinically relevant behavior and unequivocalness of diagnosis. Each *S* was correlated with every other *S* on their scores on 42 traditional Rorschach scoring categories and an inverted factor analysis computed from the matrix of intercorrelations between *S*s. None of the three orthogonal original or rotated factors were significantly related to the psychiatric grouping of the *S*s when the Rorschach response productivity of the individual *S*s was statistically eliminated. The lack of diagnostic validity of this purely psychometric approach compared to the significant validity of the clinical judgment method was attributed to several sources of information in the complete Rorschach proto-

col which are ignored by the usual scoring categories.

Received October 28, 1954.

References

1. Cattell, R. B. *Factor analysis.* New York: Harper, 1952.
2. Cattell, R. B. The three basic factor-analytic research designs—their interrelations and derivatives. *Psychol. Bull.,* 1952, **49**, 499–520.
3. Chambers, Guinevere S. An investigation of the validity of judgments based on "blind" Rorschach records. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1954.
4. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.,* 1950, **14**, 261–267.
5. Wittenborn, J. R., & Weiss, W. Patients diagnosed manic depressive psychosis—manic state. *J. consult. Psychol.,* 1952, **16**, 193–198.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Various Rorschach Indices as Discriminators of Marked and Little Conceptual Disorganization Among Schizophrenics[1]

Bernard A. Stotsky and James F. Lawrence
VA Hospital, Brockton, Massachusetts

Among the principal symptoms of schizophrenia are disorganization of thought, loss of coherence in speech, and the development of false beliefs (5). Goldstein and Sheerer (8) and Vigotsky (16) regarded the loss or impairment of abstract behavior and concept formation as the key psychological disturbance in schizophrenia. Storch (14) described schizophrenic thinking as the substitution of concrete behavior for conceptual thinking.

Rawlings (12) making use of the Yerkes-Bridges Point Scale and tests of association and imagination found marked impairment of abstraction and generalization in schizophrenia. A number of authors (2, 3, 8, 9, 16) have reported the impairment of the abstract attitude and of the ability to shift from one set to another. Cameron's (4) work did not support previous findings in that he found the "abstract capacity level" to exist among schizophrenics provided that adequate rapport and cooperation were obtained.

In the area of intelligence, Hunt and Cofer (10) have noted that schizophrenics do more poorly than nonpsychotics of similar mental age on such conceptual tests as interpreting fables, noticing absurdities, finding similarities, and on tests of associative thinking. Every author has found that the variability of performance of schizophrenics is marked, usually larger than that of any control population. In every experiment there has been a considerable amount of overlap in test scores between schizophrenics and controls. We have seen that even in an experiment (9) where very significant differences were found between schizophrenics and normals, only *26* of the *62 schizophrenics showed deficit* when comparisons were made with controls of comparable educational background.

All the investigations cited above have concerned themselves with the study of schizophrenia as a single entity and have attempted to isolate some one factor or group of factors which would differentiate the schizophrenic from the normal or from all other types of psychopathology. Only in the past few years have systematic investigations been made of differences between schizophrenics. In one study Stotsky (15) reported that schizophrenics who remit tend to show at the outset of hospitalization better intellectual and emotional control, greater clarity of the associative processes, and greater ability to concentrate than schizophrenics who remain hospitalized.

Encouraged by these findings the authors of the present study sought to investigate differences in degree of conceptual disorganization among schizophrenics. More specifically the purpose of this investigation was to examine the validity of certain Rorschach indices of conceptual disorganization on two groups of schizophrenics, one characterized by relatively little impairment of conceptual functioning, the other by a great deal of impairment.

Predictions

1. Since the proportion of good form responses in a record is regarded by Rorschach

---

[1] From VA Hospital, Brockton, Massachusetts. Rorschachs administered and scored by Murray A. Cohen, Boston University trainee.

(13) as a measure of the clarity of associative processes and by Beck (1) as an indicator of accurate perception of reality, patients manifesting little conceptual disorganization will give a greater proportion of good form responses. All responses containing a form element in them (*F, M, FC, CF, FY, YF,* and *FV*) will be scored for good form.

2. Since patients with less conceptual disorganization may be expected to retain greater ability to integrate and articulate their responses to ambiguous perceptual stimuli, these patients will show a greater proportion of genetically good location responses.[2]

3. If the number of popular responses is taken as a measure of the patient's intellectual adaptability and ability to "share in the common way of perceiving things" (13, p. 198), patients with little conceptual impairment will give a greater number of popular responses.

4. Patients with less conceptual disorganization should show greater retention of the ability to shift from one set to another. Therefore, they will show more shift from free association to inquiry for Rorschach determinants.

Procedure

Multidimensional psychiatric rating scale. Many chronic schizophrenic patients transferred to this newly opened neuropsychiatric hospital from other Veterans Administration installations were rated just prior to transfer by means of the Hospital Form of the Multidimensional Scale for Rating Psychiatric Patients (11). This form consists of 62 individual, unlabeled graphic rating scales which secure in a relatively objective and quantitative form a description of the observable behavior or readily inferable traits and common symptoms of hospitalized patients. Reliability of rating is high. The scale has discriminated severely ill from mildly ill psychotic patients for the factors of conceptual disorganization, perceptual distortion, and motor disturbances (11).

From the 62 items of the scale, 12 factors or dimensions of psychopathology have been obtained by its authors. One of these, factor *K*, "conceptual disorganization," was selected as the criterion measure for differentiating patients with marked disorganization of conceptual thinking from patients with relatively little. Factor *K* is characterized by irrelevant and incoherent speech, a stereotyped repetition of certain words and phrases and a frequent blocking in speech, a disharmony between the patient's thinking and the feeling tone shown or expressed, and a disorientation for time. When present, delusional beliefs tend to be impossible or bizarre (11, p. 6).

For each patient rated by the Multidimensional Scale the factor raw score on *K* was obtained by adding the numerical values of the ratings on the seven scale items descriptive of the factor. The items are listed below:

1. How direct and relevant are his responses to questions or to the topic discussed?

3. Are his thoughts and feelings consistent, or is there a discernible lack of harmony between them? (Reports, smiling, that he was tortured with blow torches all night.)

4. How well oriented is he as to time? Does he know (*a*) the season; (*b*) the month; (*c*) the calendar year; (*d*) the day of the week?

11. Are the elements of his speech logically consistent and connected by some idea or relationship, or do they tend to be inconsistent and disconnected? Rate what is most representative during the interview.

28. Does he repeat certain words or phrases in a meaningless, stereotyped, or mechanical fashion?

29. Is his speech irregularly interrupted, halted, or blocked for varying periods of time because he is not able to put the idea he has in words? (Do not rate stuttering, stammering, or muteness here.)

38. Is there evidence of false ideas or beliefs? If present, are these ideas or beliefs (*a*) sufficiently plausible as to be accepted by a normal person uninformed as to the facts; (*b*) implausible but not impossible; (*c*) impossible or bizarre (e.g., mind controlled by neighbor's radio waves, heart removed, or dead)?

Items 1, 3, 11, 28, 29, and 38 were rated on a four-point scale, item 4 on a five-point scale. Patients obtaining a score of 10 or below for *K* (10 was the mean score for *K* of the reference population drawn from five VA hospitals) were considered to show relatively low conceptual disorganization. Patients with scores of 16 or greater for *K* (highest 17% of reference population) were considered to show high conceptual disorganization.

---

[2] The scoring of the quality of location (whole, common detail, and rare detail) responses from a genetic point of view was accomplished by means of Friedman's scoring criteria (6).

Sampling. Patients were randomly selected from those scoring 10 or less and 16 or greater on *K*, and a Rorschach was administered individually to each testable patient within 60 days of admission to this hospital. The goal was to obtain 10 Rorschach records for each group. Of the first 15 patients approached in the low group, 10 cooperated sufficiently to give scorable Rorschachs with more than six responses. In the high group only 10 patients out of 23 were testable to the extent of meeting the criterion of seven responses to the 10 cards.

Although all patients had taken Rorschachs within 60 days of admission, it was considered desirable to rerate the patients on the seven criterion items of the scale to see whether any changes had taken place. The *K* scores for all 10 lows remained constant. For seven of the highs the scores remained constant. The scores for two increased slightly. The tenth high, however, had shown such marked improvement that his *K* score fell below 16, necessitating his elimination from the group. As finally constituted, the total sample consisted of 10 low scorers and 9 high scorers on *K*.

All subjects were white, male veterans below the age of 60. Comparison of the two groups for age, length of VA hospitalization, education, level of intelligence, schizophrenic subgroup, premorbid occupation, degree of rated disability, marital status, and religion showed no significant differences. In no case was there any evidence reported of cerebral lesion or neurological deficit. All patients had confirmed histories of schizophrenic symptomatology, accompanied by long periods of hospitalization.

Rorschach scoring. The Rorschach protocols were scored for both free association and free association plus inquiry, using Beck's system for scoring determinants, good and poor form, and popular responses. For location of responses the proportion of all good or mediocre, *W*, *D*, and *Dd* responses over total responses was used to compare the groups rather than considering each of the location categories separately. Thus, for *W*, *D*, and *Dd*, Friedman's (6) plus-plus, plus, and mediocre scores were categorized as good, and vague, confabulatory, and minus responses were categorized as poor. The degree of agreement between two clinicians scoring location by means of Friedman's criteria was greater than 85%, which was considered adequate for this study.

Prior to the analysis of findings both groups were compared for total number of responses (*R*). No significant difference was found between the mean *R* of 17.8 for the low group and of 16.3 for the high group.

Findings

1. *Prediction for good form.* The prediction for good form was sustained at the .001 level of significance with the low group giving a significantly greater proportion of good form responses. There was actually no overlap between the groups inasmuch as the highest score for the highs was 70 and the lowest for the lows was 77.[3]

Table 1

Comparison of High and Low Patients for Rorschach Indices of Conceptual Disorganization

Rorschach variable	Highs		Lows				
	Mean	Sigma	Mean	Sigma	t	χ^2	p
Good form per cent	47.0	17.1	86.4	6.0		15.4	.001
Good location per cent	38.1	23.2	82.6	8.2		15.4	.001
Populars	2.7	1.8	5.1	1.2	3.3		.01
Total determinant shift	1.4	1.2	3.1	1.2	2.4		.03

2. *Prediction for location.* The difference between the low and high groups was significant at the .001 level in the predicted direction for location. Here also there was no overlap on scores between the two groups, the highest score for the highs being 63, the lowest for the lows, 71.

3. *Prediction of populars.* The prediction with regard to *P* was supported at the .01 level of significance with the low group giving almost twice as many popular responses as the high group.

---

[3] For Predictions 1 and 2 it was necessary to use chi square rather than *t* in the analysis of the data since the variances for the high group were significantly greater than those for the low group. For good form per cent the cutting point was 75, for good location per cent the cutting point was 70.

4. *Prediction for total determinant shift.* The low group showed significantly greater shift ($p = .03$) for determinants from free association to inquiry. This confirms the prediction that such a group is better able to shift from one set to another and is more responsive to a change in stimulus.

Discussion

All four predictions concerning the Rorschach indices of conceptual disorganization were confirmed below the .05 level of significance. Of greater importance was the fact that for two of these indices, good form per cent and good location per cent, there was no overlap between the two groups. Tentatively these findings suggest that low scores for these two Rorschach variables may be taken to indicate conceptual disorganization among schizophrenic patients.

The findings for form and populars support the notion that the schizophrenics low on conceptual disorganization are able to associate accurately to the blots and to share in the common way of perceiving things. The superior performance of the lows on the location categories suggests that these patients not only retain their ability to integrate their responses to the plates, but that they may even tend to approach the level of response of neurotics and normals on these variables.

While the difference between the two groups for determinant shift is significant, it is noteworthy that neither group shifts a great deal. Gibby and Stotsky (7) in comparing psychotic and psychoneurotic patients found that the average amount of shift for the psychotic group was 3.15 while that for the neurotic group was greater, 5.15. Thus in the present study the psychotic patients with low conceptual disorganization show roughly the same amount of shift as the psychotics studied at a VA general medical and surgical hospital (7). The average shift for the high and low groups combined is much lower, roughly 2.3. The smaller amount of shift for these patients as compared with the psychotics in the earlier study may be due to their being more chronically schizophrenic and as a result of their illness probably less responsive to the shift in stimulus from free association to inquiry.

Of what significance are the present findings with regard to schizophrenia in general? The definite differences, on both the rating scale and the Rorschach indices, between the two groups, both of which consisted of chronic schizophrenic patients, raise the question as to the advisability of regarding all schizophrenics as individuals incapable of higher level thinking such as is found in normals. The present findings indicate that there are many schizophrenic patients who are able to shift from one set to another and who can integrate perceptual stimuli in a manner similar to that of nonschizophrenic subjects. Among such patients the schizophrenic process does not manifest itself by marked conceptual disorganization. In fact they show minimal conceptual impairment on the measures employed in this study.

Summary

1. By means of the Multidimensional Psychiatric Rating Scale, 10 chronic schizophrenic patients showing little conceptual disorganization and 9 chronic schizophrenic patients showing marked conceptual disorganization were selected and compared on four Rorschach indices of conceptual disorganization: good form per cent, good location per cent, populars, and determination shift.

2. It was predicted that patients low on conceptual disorganization would show a greater proportion of good form and good location responses, a greater number of populars, and more determinant shift from free association to inquiry than patients high on conceptual disorganization. All predictions were confirmed at significant levels of confidence.

3. The findings were discussed with regard to their application to the consideration of schizophrenia as a single entity.

Received August 19, 1954.

References

1. Beck, S. J. *Rorschach's test.* Vol. II. *A variety of personality pictures.* New York: Grune & Stratton, 1947.
2. Bolles, Marjorie M., & Goldstein, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, **12**, 42–66.

3. Bychowski, G. Certain problems of schizophrenia in the light of cerebral pathology. *J. nerv. ment. Dis.*, 1935, **81**, 280–298.
4. Cameron, N. Schizophrenic thinking in a problem-solving situation. *J. ment. Sci.*, 1939, **85**, 1012–1035.
5. Cameron, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 861–921.
6. Friedman, H. Perceptual regression in schizophrenia. An hypothesis suggested by the use of the Rorschach test. *J. proj. Tech.*, 1953, **17**, 171–185.
7. Gibby, R. G., & Stotsky, B. A. Determinant shift of psychoneurotic and psychotic patients. *J. consult. Psychol.*, 1954, **18**, 267–270.
8. Goldstein, K., & Sheerer, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
9. Hanfmann, Eugenia, & Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, No. 67.
10. Hunt, J. McV., & Cofer, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 971–1032.
11. Lorr, M. Multidimensional scales for rating psychiatric patients. Hospital form. *Vet. Adm. tech. Bull.* 10–507, Nov. 16, 1953.
12. Rawlings, E. The intellectual status of patients with paranoid dementia praecox: its relation to organic brain changes. *Arch. Neurol. Psychiat.*, 1921, **5**, 283–295.
13. Rorschach, H. *Psychodiagnostics*. Berne: Hans Huber, 1942.
14. Storch, A. The primitive archaic forms of inner experience and thought in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1924, No. 36.
15. Stotsky, B. A. A comparison of remitting and nonremitting schizophrenics on psychological tests. *J. abnorm. soc. Psychol.*, 1951, **47**, 489–496.
16. Vigotsky, I. S. Thought in schizophrenia. *Arch. Neurol. Psychiat.*, 1934, **31**, 1063–1077.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

The Separation Capacity of the Rorschach

Raymond J. Corsini,
University of Chicago

Winfield E. Severson, Thomas E. Tunney, and Harold F. Uehling
Wisconsin Department of Public Welfare

The Rorschach may be used to rate individuals on single or multiple dimensions with respect to relative or absolute standards. Although the literature on the Rorschach has been concerned mainly with the adequacy of clinical descriptions of individuals, in some situations, as in a prison, it is often of greater practical importance to rate people with respect to one over-all quality, such as "adjustment." For classification boards and parole committees the important question is "How is he?" rather than "What is he?" Prison officials want to know whether an individual placed in a particular situation of trust will be able to meet it effectively, and they are not too concerned with why or how he will succeed. They are interested in meaningful predictive statements and not in understanding individual dynamics.

In either case, the validity of the Rorschach is at issue. The difficulties inherent in the validation of projective techniques have been ably summarized by Macfarlane and Tuddenham (2) and need not be reviewed again. One kind of validation may be made by testing for criterion discrimination, that is the capacity of the Rorschach to separate individuals known or believed to differ with respect to some criterion quality. However, to eliminate other cues, such as will occur in face-to-face contact between the examiner and the subject, it is necessary to have Rorschachs evaluated blindly. This is particularly true in this investigation, since half of the subjects were prisoners, the other half were prison guards.

The Problem

In a prior paper (1) it was reported that, when 50 "normal" Rorschach and 50 "abnormal" Rorschachs were evaluated by means of a check list, the rank-ordered ratings placed 29 normals in the top 50 rather than the chance expectancy of 25. The degree of separation was not quite significant at the 5 per cent level. The question arises of the capacity of psychologists to separate such protocols by clinical judgment in contrast to the mechanical procedure of the check list.

Method

The same 100 protocols used in the prior investigation were used in this study. The 50 normal protocols were of newly hired prison guards who took the Rorschach as part of the routine of indoctrination into the penal service. They had passed civil service investigations of their character, had the equivalent of high school education, and had passed two interviews designed, among other purposes, to evaluate their emotional stability. There was no reason to believe that they were in any way different with respect to normality of personality from an unselected group of the same age. If anything, they may have been superior, as a result of the screening. The 50 abnormal protocols were of prisoners who had committed serious felony crimes and who had been referred to the psychologists for projective testing by the psychiatrist because in his opinion they showed signs of serious personality deviations. About 10 per cent of prisoners in this particular institution are referred for further intensive testing which usually includes the Rorschach.

Each of the guard protocols was matched with that of a prisoner. Ages were balanced within three years; only inmates who had ratings of average intelligence, and who came from lower middle social classes, were used. All subjects were white.

The protocols were assembled with a summary sheet. All identifications were removed. The protocols were then given to three psychologists with instructions to rank them from 1 to 100 with respect to the most probable adjustment of the subjects. Psychologists A and B were asked to take as much time as they considered necessary to make a good decision. Psychologist C was asked to make "snap" decisions. Psychologist D, who had administered all the Rorschachs, was asked to rank the summaries of his reports from 1 to 100 in terms of favorableness with respect to adjustment.

Results

Psychologist A averaged 7.2 minutes per judgment; Psychologist B, 6.0 minutes; and Psychologist C, 1.1 minute. Time was not recorded for Psychologist D. The intercorrelations of the rankings between the psychologists were very low, in terms of usual psychometric expectations. The rankings of A and B, A and C, and B and C, correlated only .22, .35, and .45, respectively. The judgments of these three psychologists, respectively, against the ratings of the report summaries prepared by Psychologist D, correlated .33, .58, and .50.

By chance it might be expected that 25 of the normal protocols might be placed in the top half of the rankings. It had been found that through using the Davidson Scale 29 were placed in the top half. Psychologists A and B placed, respectively, 34 and 42. Both exceeded the 1 per cent level of significance. Psychologist C, who worked very rapidly at the rate of about one judgment per minute, exceeded the Davidson Scale by placing 30 of the guard protocols rather than the chance expectancy of 25, thereby just meeting the 5 per cent level of significance. All three of the psychologists, who used their clinical judgment, were better than a check list for the purpose of separating this population of Rorschach protocols with respect to "normality." Psychologist D, who had worked from his own summary reports, put 40 of the guards in the top 50. Psychologist D's results, however, could be seriously contaminated because he saw the examinees in face-to-face situations and could have been influenced by clues not obtained from the Rorschach. The placements of the four psychologists in terms of quartiles is shown in Table 1.

Table 1

Number of Guard Protocols in Each of the Four Quartiles of Ratings

	Quartiles			
Psychologist	I	II	III	IV
A	20	14	10	6
B	23	19	6	2
C	18	12	12	8
D	24	16	8	2

Summary

The most surprising result of this investigation is the very low order of reliability, in terms of usual psychometric expectations, found in the rankings of the Rorschach protocols by the psychologists with respect to normality. Of course, no information exists as to the range of "normality-abnormality" of the subjects, but presumably there was no attenuation due to restriction of range.

It would appear that adequacy of judgments is a function of time spent on the judging process. The three psychologists were considered equal in their abilities, but the psychologist who made the "snap" judgments did more poorly than the other two who worked at what they considered optimal speed. The best score of all, obtained by Psychologist B, a new employee of the Department of Public Welfare, was obtained despite the fact that he had not known that the protocols were of guards and prisoners.

The Davidson Rorschach Adjustment Scale selected 58 per cent of normals, the psychologists, 60, 68, and 82 per cent from the protocols, and the psychologist working from his own reports, 80 per cent.

It may be concluded that psychologists' clinical judgments of Rorschachs, while quite deficient in agreement with one another, are nevertheless superior to the Davidson Rorschach Adjustment Scale for the purpose of separating two groups of adults who differ with respect to over-all social adjustment.

Received September 27, 1954.

References

1. Corsini, R. J., & Uehling, H. F. A cross validation of Davidson's Rorschach Adjustment Scale. *J. consult. Psychol.*, 1954, **18**, 277–279.
2. Macfarlane, Jean W., & Tuddenham, R. D. Problems in the validation of projective techniques. In H. H. Anderson & Gladys Y. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 26–42.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

The Effect of the Psychotherapist's Personal Analysis Upon His Techniques[1]

Hans H. Strupp
Washington, D. C.[2]

Two previous publications have presented objective evidence on (*a*) the techniques of Rogerian and psychoanalytically oriented psychotherapists (5), and (*b*) the verbal behavior of psychiatrists, psychologists, and psychiatric social workers of varying levels of experience (6). These articles also set forth the rationale for focusing attention upon the psychotherapist and for studying systematically the concomitants of his techniques. The present paper is an extension of the previous inquiry, and attempts to answer the question: What is the effect of the therapist's personal analysis upon his therapeutic techniques?

Since the time of Freud no training requirement for the psychotherapist has been considered more important than his personal analysis, and the curricula of psychoanalytic training institutes have traditionally been built around this aspect of the candidate's training. In view of the voluminous literature on the subject, little need be said about the arguments commonly advanced in favor of this training requirement. Fromm-Reichmann's summary, while perhaps unduly restrictive, is representative of current thinking: "And so it is that, because of the interrelatedness between the psychiatrist's and the patient's interpersonal processes and because of the interpersonal character of the psychotherapeutic process itself, any attempt at intensive psychotherapy is fraught with danger, hence unacceptable, where not preceded by the future psychiatrist's personal analysis" (*2*, p. *42*).

If this position is correct, the experience of personal analysis should have a demonstrable effect upon the character of the therapist's operations. Even if present-day methods for objectively examining such an elusive variable are crude, there must be systematic differences between the techniques of therapists whose training has included personal analysis and those whose training has not. After the exploration of this general hypothesis, the problem will be pursued with reference to a few salient therapeutic situations.

Three Kinds of Patient Maneuvers

Suicide threats. A suicide threat is one kind of power operation by which the patient attempts to bring a significant person under his control. The reaction is, of course, the culmination of a long series of frustrations, rebuffs, and injuries which have been inflicted in reality or fantasy upon the victim, and constitutes a grandiose attempt at revenge by turning the tables upon the perceived aggressor. Without attempting to give the dynamics more than cursory treatment, it may be recognized that a suicide threat contains as its essential ingredients: (*a*) profound disappointment, frustration, and suffering; and (*b*) pervasive reactions of rage, aggression, and hostility.

Operationally, self-destructive threats expressed by a patient may serve to evoke the therapist's pity and commiseration, they may be designed to perpetuate a dependent relationship, or they may represent a maneuver to dominate and to make the therapist responsible for the patient's actions.

---

[1] This article is the third in a series of investigations concerned with the objective study of psychotherapists' verbal operations, and represents one phase of the author's doctoral dissertation submitted to the Graduate Council of the George Washington University (4). Part of this material was presented at the annual meeting of the American Psychological Association in New York, September 1954. The writer wishes to express his appreciation to Dr. Dorothy E. Green for valuable statistical advice.

[2] While carrying out this work, the author held positions successively with the Department of the Air Force and the Department of the Army. Neither organization has any connection with this research.

With regard to recommended therapeutic approaches, Fromm-Reichmann states: "A thorough investigation of the validity of the causes for the patient's discouragement, unhappiness, or despair must be the starting point of every therapeutic approach to suicide, or suicidal attempts and fantasies" (2, p. 198). By the same token, reassurance, pity, etc. appear to be ill-advised: "If the therapist offers encouragement to patients' parataxic expectations by actually falling into the role of practical adviser, he retards the process of insight into the immature character of such expectations; hence he retards the process of resolving them" (2, pp. 208–209).

The following hypothesis tests whether therapists do what they preach:

Hypothesis 1. When confronted with suicide threats, psychotherapists include in their responses (*a*) a relatively large number of explorations; (*b*) a relatively small number of responses that convey reassurance; and (*c*) a relatively small number of interpretations.

If the therapist who has undergone personal analysis is more clearly aware of the dynamics of the situation, Hypothesis 1 may be extended by stating that the predicted trend is more pronounced for therapists whose training has included personal therapy.

Transference reactions. In its special psychotherapeutic meaning, transference refers to all processes whereby the patient re-enacts in the therapeutic situation unresolved interpersonal conflicts, and casts the therapist into the role of the significant adult who figured prominently in the earlier experience. Transference constitutes the cornerstone of all modern conceptions of psychoanalytic therapy and is increasingly utilized by other forms of therapy as well. This emphasis also implies that a significant segment of the therapeutic work consists in clarifying the doctor-patient relationship and concomitantly the distortions which complicate the patient's living.

The giving of interpretations has to be carefully planned and timing is all-important. Interpretations which are offered prematurely, that is, before the establishment of a relatively durable doctor-patient relationship and relaxation of the patient's emotional defenses, are useless if not positively harmful. Still, transference reactions manifested by the patient at almost any stage of the treatment are likely to be dealt with by means of interpretations, particularly by analyzed therapists who themselves have been exposed to the psychoanalytic maxim of giving primary attention to transference phenomena. The second hypothesis is:

Hypothesis 2. When confronted with transference reactions, psychotherapists employ a relatively large number of interpretations. Analyzed therapists exhibit this tendency to a greater extent.

Schizoid productions. The third hypothesis deals with the productions of a single patient who appears seriously disturbed, and who gives the impression of being on the verge of a schizophrenic break. It appears plausible that such a patient would be treated differently early in therapy than one who is less disturbed, anxious, and confused. For one thing, the therapeutic effort obviously must be directed at bringing about a diminution rather than an increase in anxiety; for another, the therapist must attempt to obtain a fuller picture of the underlying dynamics. Since the technique of silence is often used to create "a sort of vacuum [which] pushes the patient into taking the initiative in the relationship" (7, p. 221), it increases of necessity the patient's anxiety, for which reason it would be contraindicated with a very anxious patient. As for the second objective, a fair number of exploratory questions may be anticipated. On the assumption that the analyzed therapist will be more sensitive to the requirements of the situation, Hypothesis 3 may be stated:

Hypothesis 3. When confronted with the productions of a seriously disturbed, near-psychotic patient, psychotherapists employ (*a*) a relatively small number of silent responses, and (*b*) a relatively large number of explorations. Analyzed therapists follow this trend in more clear-cut fashion.

Method, Procedure, and Subjects

A fuller description has been given in the earlier article (6). For the present, it suffices to reiterate that samples of psychotherapists' verbal behavior were elicited by presenting a series of 27 cards containing short paragraphs of patient statements selected from published therapeutic interviews. Included in the series, which was considered a fair cross section of verbalizations heard from neurotic patients in early interviews, were four suicide threats, six transference reactions (including an insistent request for direct advice, denials of the need for therapy, and examples of competitiveness and open hostility), and six complaint statements of a seriously disturbed near-psychotic patient.[3]

Table 1

The Sample by Professional Affiliation, Experience Level, and Personal Analysis

Group	Personal analysis			No personal analysis			
	Exp.	In-exp.	Sub-total	Exp.	In-exp.	Sub-total	Total
Psychiatrists	14	3	17	1	7	8	25
Psychologists	3	3	6	0	1	1	7
Psychiatric social workers	6	1	7	1	1	2	9
Total	23	7	30	2	9	11	41

Table 2

Brief Description of Scoring Categories

Category number	Core meaning
1	Reassurance
2	Not applicable*
3	Silence (passive acceptance)
4	Structuring
5	Interpretation
6	Reflection of feeling
7	Factual questions
8	Exploration
9	Not applicable*
10	Passive rejection
11	Not applicable*
12	Antagonism

* Probably a function of the experimental conditions under which responses were elicited.

Subjects were the same 25 psychiatrists, 7 psychologists, and 9 psychiatric social workers who participated in the previous research (6). All were psychoanalytically oriented. Length of experience in psychotherapy ranged from 1 to 12 years, with a median of 5 years. Practically all of the experienced therapists (5 years and above), and about half of the inexperienced therapists (less than 5 years) had undergone personal analysis as part of their training. These data are shown in Table 1. While no systematic information was obtained concerning the length of a respondent's personal analysis, it is safe to assume that for the most part it had been fairly extensive—certainly beyond one year. A few therapists who had only recently started their analysis were included in the unanalyzed group.

Results

Quantification of therapists' responses was accomplished by means of Bales' (1) system of interaction process analysis. A total of 1,609 score units was categorized by this method. Rater agreement was tested on a stratified random sample of 370 score units (10 cases) and found to be 78 per cent. A brief description of the scoring categories is given in Table 2.[4]

Figure 1 presents the response profiles of analyzed and unanalyzed therapists, regardless of professional affiliation. The statistical analyses test the significance of percentage differences in a particular response category; *t* values significant at the .01 or .05 level are indicated in the figure. The findings may be summarized as follows:

1. The two response profiles show considerable overlap in several of the major categories; nevertheless, a chi square computed on an over-all basis is significant beyond the .001 level of confidence.

2. A pronounced discrepancy occurs in Category 3, denoting that analyzed therapists give significantly fewer silent (passive accept-

---

[3] Adapted from Rogers' case "Miss Tilden," reported in Snyder (3); by permission of Professor Snyder.

[4] For a fuller description of the scoring categories, see the earlier publications (5, 6).

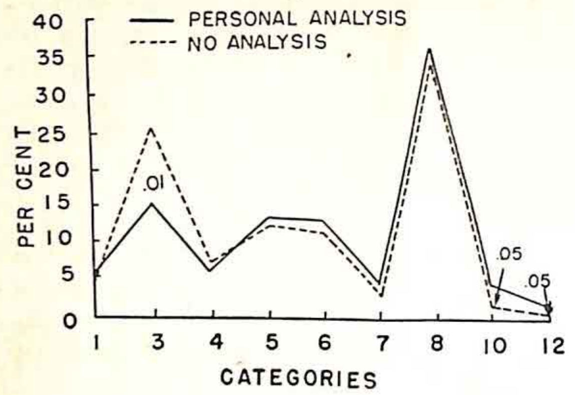

Fig. 1. Response distributions of analyzed and unanalyzed therapists (personal analysis, $n = 1{,}173$; no personal analysis, $n = 436$).

ance) responses, and conversely that unanalyzed therapists are more passive.[5]

3. Small, but statistically significant differences in Categories 10 and 12 (passive rejection and antagonism), while possibly quite real, are based upon small frequencies and should be regarded as tentative.

Since the mean length of experience of analyzed therapists was significantly greater than that of unanalyzed therapists ($F = 6.33$, $p < .05$), there was reason to question whether the significant difference between analyzed and unanalyzed therapists with respect to the number of silent responses was attributable to personal analysis, or whether it was rather a function of length of experience. The analysis of covariance technique was employed to determine whether the difference between the number of silent responses by analyzed and unanalyzed therapists was significant when the effect of experience was controlled. The obtained F value of 9.39 ($p < .01$) indicates that, as far as the present data are concerned, the difference in silent responses between analyzed and unanalyzed therapists is attributable to the variable of personal analysis. This result was to be expected since the correlation between length of experience and number of silent responses is .12, which is not significantly different from zero.

To test the three specific hypotheses, therapists' responses to each subseries (suicide threats, transference reactions, and schizoid productions) were compared with responses to all other statements. These operations were performed separately for analyzed and for unanalyzed therapists. Subseries responses were, of course, always excluded from the over-all frequencies.

Suicide threats. Figure 2 shows the response profiles of analyzed and unanalyzed therapists to suicide threats and to all other statements. The following differences may be observed:

1. Unanalyzed therapists give a *larger* number of reassuring responses (Category 1) to suicide threats; for analyzed practitioners the trend is in the same direction ($t = 1.95$, $p < .06$).

2. Unanalyzed therapists disclose a significant decrease in reflective responses (Category 6).

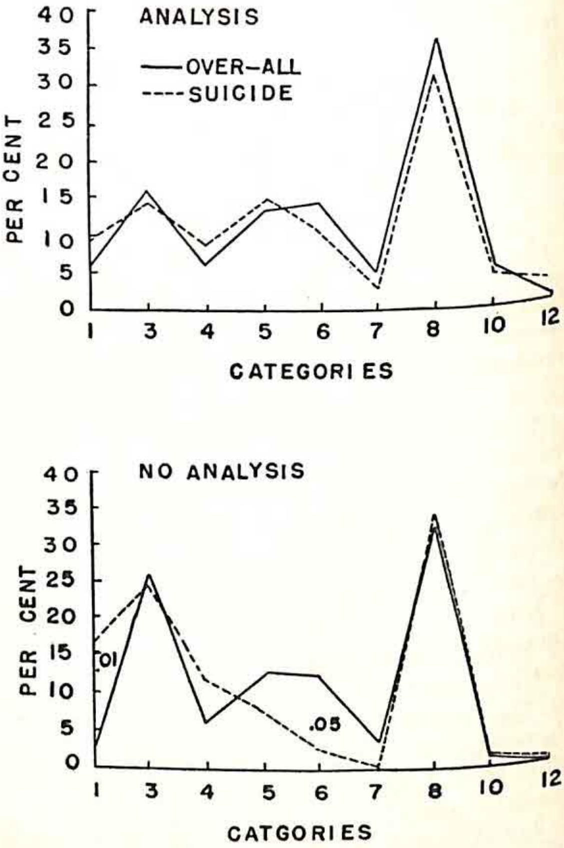

Fig. 2. Response distributions of analyzed and unanalyzed therapists to suicide threats and to all other statements (personal analysis: suicide, $n = 182$; over-all, $n = 991$; no personal analysis: suicide, $n = 64$; over-all, $n = 372$).

---

[5] The findings reported for Rogerian therapists (5) corroborate this result.

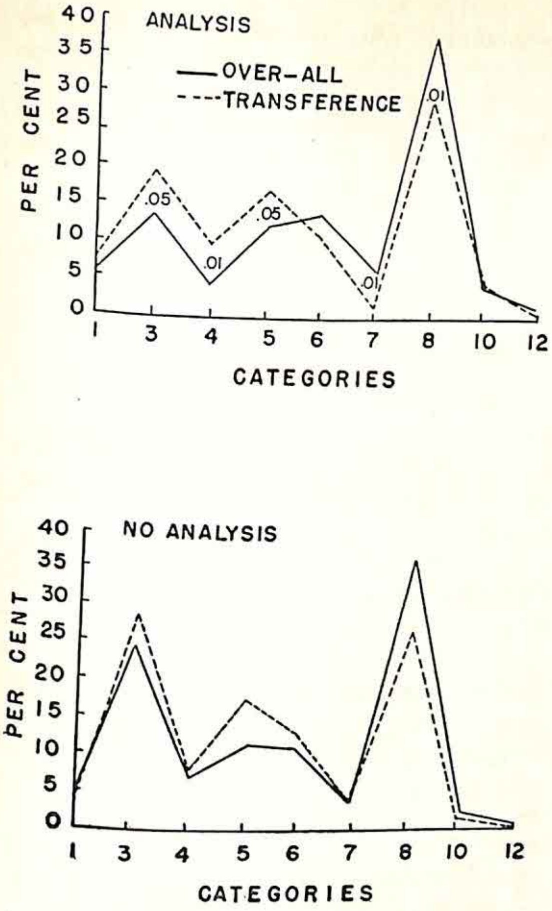

Fig. 3. Response distributions of analyzed and unanalyzed therapists to transference reactions and to all other statements (personal analysis: transference, $n = 273$; over-all, $n = 900$; no personal analysis: transference, $n = 99$; over-all, $n = 337$).

3. Both groups show a slight preference for structuring responses (Category 4) when dealing with suicide threats; however, neither difference is significant at the .05 level (analyzed group: $t = 1.63$, $p < .11$; unanalyzed group: $t = 1.94$, $p < .06$).

None of the above outcomes is in accordance with Hypothesis 1; in fact, the increment in reassuring responses is in opposition to the prediction, except that unanalyzed therapists appear to give a relatively larger number of reassuring responses than do analyzed workers.

Transference reactions. Figure 3 compares the response profiles of analyzed and unanalyzed therapists to transference reactions and to all other statements. The following observations may be made:

1. Both groups show an increase in interpretive responses (Category 5), but the difference is statistically reliable only in the case of the analyzed therapists.

2. Both analyzed and unanalyzed practitioners tend to give a larger proportion of silent responses (Category 3), but again the difference is significant only for the former group. It should be recalled here that this category revealed a significant discrepancy in the comparison of over-all distributions, unanalyzed workers exceeding the analyzed therapists.

3. Increments in structuring (Category 4) and decrements in exploratory responses (Category 8) are apparent for the analyzed group. These trends appear to be paralleled by the unanalyzed group, but here the differ-

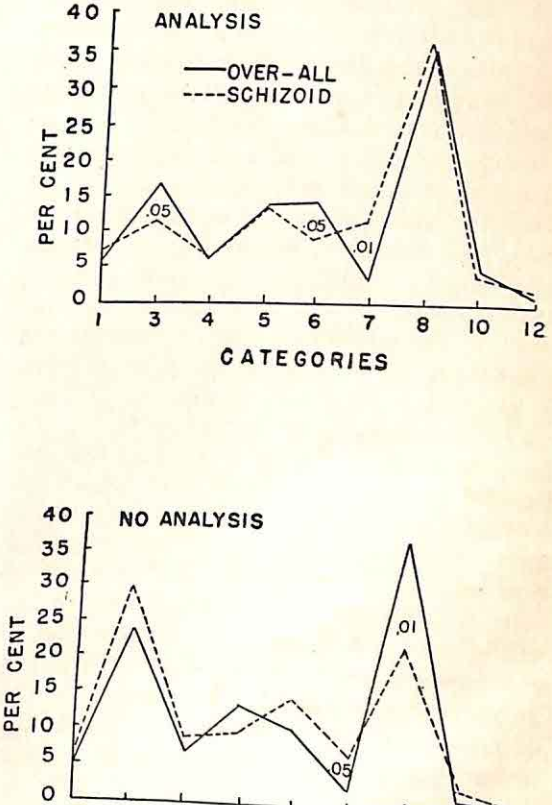

Fig. 4. Response distributions of analyzed and unanalyzed therapists to schizoid productions and to all other statements (personal analysis: schizoid productions, $n = 274$; over-all, $n = 899$; no personal analysis: schizoid productions, $n = 94$; over-all, $n = 342$).

ences are not significant, probably because the absolute numerical frequencies are considerably smaller. The difference in Category 7 is inconsistent and may be due to chance.

Hypothesis 2, which predicted an increase of interpretive responses, is thus partially confirmed; however, the verbal behavior of analyzed therapists seems to follow the same trend as that of their unanalyzed confreres. Collateral statistical treatments indicated that psychologists and social workers do *not* contribute to this shift.

Schizoid productions. Figure 4, which presents the by now familiar comparisons for schizoid productions, leads to the following interpretations:

1. Analyzed therapists' responses are characterized by a marked *decrease* in silent responses (significant at the .05 level), whereas unanalyzed therapists reveal an *increase* (not statistically significant).
2. Analyzed therapists use a proportionate number of explorations with this patient; unanalyzed workers, on the other hand, give significantly fewer exploratory responses.
3. The differences in Category 7 (direct questions) are traceable to an artifact in the items of the subseries. Another inconsistent difference in Category 6 may be due to chance.

Hypothesis 3 is partially confirmed by these results, particularly with respect to the decrease in silent responses for analyzed workers. If it is recalled that analyzed therapists were shown to be more active in the over-all comparisons, it follows that the gap between the analyzed and the unanalyzed groups is widened when responses to this patient occupy the focus of the statistical treatment. Exploratory responses, while undergoing no shift in the case of analyzed therapists, reveal a pronounced decrease in the unanalyzed group. This result is likewise in the predicted direction.

Discussion

In interpreting the results of this study certain limitations must be kept in mind:

1. Therapists' responses were secured by means of an experimental model whose validity remains to be tested. It is thus not known to what extent the therapist's behavior in the experimental situation coincides with his behavior vis-à-vis a patient.
2. The experimental statements were brief and out of context; there was a minimum of background information on each patient; and, of course, the all-important therapist-patient relationship was lacking.
3. The three specific hypotheses were tested only in an approximate manner, and the tests are altogether relative since the reference distributions were composed of heterogeneous elements whose representativeness is not known. The number of suicide, transference, and schizoid items, moreover, was small so that reliable differences were difficult to demonstrate.

Within these limitations, the present investigation provides objective evidence on a problem about which much has been written but on which no quantitative data have yet been adduced.

The finding that analyzed therapists tend to be more active than their unanalyzed colleagues is a provocative one, but it runs counter to an intuitive hunch to the effect that the experience of personal analysis enables the therapist to maintain distance from the patient's maneuvers, and consequently that he feels less compelled to respond to any and all of the patient's communications. Rather, the present results may signify that the analyzed therapist is more skilled or more willing to formulate an immediate response. An alternative interpretation might be that the unanalyzed therapist is more passive when confronted with a series of patient statements in an experimental setting, but that this trend would be reversed in actual therapy. He might be more wary, more hesitant, and more unsure of himself and of his therapeutic procedures when called upon to give an account of himself, so to speak, to an outsider. It must be remembered here that length of experience per se *does not* account for the difference which, by this evidence, seems to be due to personal analysis or some related variable.

With respect to suicide threats, it is seen that contrary to recommendations made in the literature, *all* therapists tend to emphasize reassuring responses. At the same time, we may observe a consistent but statistically not significant increase in structuring responses which attempt a clarification of the patient's and the therapist's roles in the treatment situation. Such responses emphasize the reality aspects in opposition to the parataxic character of the patient's perceptions. The latter finding, however, must be regarded as very tentative. Also, it should not be forgotten that exploratory responses, while evidently not subject to shift in this instance, account for about a third of all therapist responses.

In accordance with psychoanalytic theory, therapists appear to single out transference phenomena for interpretive attention, although the experience of personal analysis seems to have little effect on the therapist's verbal behavior in this case. As has been shown earlier (6), a predilection for interpretations is a distinguishing feature of all psychiatrists. The data would indicate that transference reactions act as a signal for the therapist to interpret, to be silent, or to define the therapeutic situation. Also, he seems to be less inclined to ask the patient for an elaboration of his feelings. However, since the findings for the smaller unanalyzed group lack conclusiveness, it would be less than cautious to attribute the observed differences to the variable of personal analysis.

The data presented for the productions of a very anxious and near-psychotic patient point to differential handling by analyzed and unanalyzed therapists, and tend to confirm the third hypothesis. Silent responses appear to be sensitive to this kind of stimulus material such that analyzed therapists become more active (the evidence is not conclusive for unanalyzed practitioners). The decline in exploratory responses by unanalyzed therapists may be the complement of this trend. In the context of the present results, this may perhaps be regarded as tentative evidence for the analyzed therapist's greater sensitivity to the demands of the therapeutic situation; and if one meaning underlying silent responses is the attitude of "playing it safe," there may be here an indication of the analyzed therapist's greater versatility and readiness for verbal participation. However, this extrapolation is clearly a speculation.

In conclusion, this study has presented incomplete but certainly provocative evidence concerning differences in the therapeutic behavior of analyzed and unanalyzed psychotherapists. Contrary to prediction, personal analysis seems to lead to greater rather than to diminished activity on the therapist's part. The implications as well as the generalizability of these findings remain to be explored further, but there can be no doubt that the problem is one of the first magnitude—for theory, for practice, and for training.

Summary

This is the third and concluding article in a series of investigations to elucidate the psychotherapist's contribution to the treatment situation. It studies the effect of the therapist's personal analysis with reference to a series of patient communications and to (*a*) suicide threats, (*b*) transference reactions, and (*c*) schizoid productions. Therapeutic responses were secured from 25 psychiatrists, 7 psychologists, and 9 psychiatric social workers of varying degrees of professional experience by presenting a series of 27 patient statements extracted from actual therapeutic interviews. A total of 30 therapists had undergone personal analysis as part of their training. Therapists' responses numbering 1,609 were categorized by Bales' system of interaction process analysis. Average rater agreement was 78 per cent. The major results may be summarized as follows:

1. Compared with unanalyzed therapists, analyzed practitioners tend to be more active, as evidenced by a significantly smaller number of silent responses.

2. *Suicide threats* evoke an increased number of reassuring responses from both therapist groups.

3. In dealing with *transference phenomena*, analyzed therapists tend to prefer interpretations, silence, and structuring responses. The results for the unanalyzed group are inconclusive.

4. *Schizoid productions* of a seriously disturbed patient appear to induce a smaller number of silent responses in analyzed thera-

pists and a smaller number of exploratory responses in unanalyzed therapists. Since the foregoing results are significant only for one group, they must be considered tentative.

Within the limitations of this study, personal analysis has a demonstrable effect on the therapist's verbal behavior. It was shown that this effect is independent of the therapist's level of experience.

This series of investigations has indicated the feasibility of studying objectively some aspects of the psychotherapist's techniques. There is reason to believe that this focus on the therapist's contribution to the treatment situation is of theoretical and practical value for advancing our knowledge of the process of psychotherapy.

Received October 8, 1954.

References

1. Bales, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley Press, 1950.
2. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univer. of Chicago Press, 1950.
3. Snyder, W. U. (Ed.) *Casebook of non-directive counseling.* Boston: Houghton Mifflin, 1947.
4. Strupp, H. H. An empirical study of certain psychotherapeutic operations: an exploration of the verbal response techniques of psychiatrists, psychologists, and psychiatric social workers. Unpublished doctor's dissertation, George Washington Univer., 1954.
5. Strupp, H. H. An objective comparison of Rogerian and psychoanalytic techniques. *J. consult. Psychol.,* 1955, **19,** 1–7.
6. Strupp, H. H. Psychotherapeutic technique, professional affiliation, and experience level. *J. consult. Psychol.,* 1955, **19,** 97–102.
7. Whitaker, C. A., & Malone, T. P. *The roots of psychotherapy.* Philadelphia: Blakiston, 1953.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Changes in the Self Concept Without Psychotherapy[1]

Donald M. Taylor
Vanderbilt University [2]

Psychologists approaching the problem of evaluation of psychotherapy have reported that a number of alterations in the self concept seem to accompany "improvement" as rated by the therapist. Studies by Raimy (5, 6), Rogers (7, 8, 9), Hartley (3), Dymond (2), and others indicate that the following changes are characteristic:

1. The self concept becomes more positive.
2. It becomes more congruent with the self-ideal.
3. It becomes more self-consistent.

These, then, might be considered possible criteria for success in counseling. One difficulty has been the lack of information on self-concept changes which take place without therapy. Dymond (2) has summarized data on one control group of 15 noncounseled subjects at the University of Chicago Counseling Center which indicate no significant change in positive-negative adjustment score over a period of several months.

Problem

To obtain data on changes in the self concept for persons not receiving counseling, a self-concept scale was devised and utilized in replicated *Q*-sort descriptions of self by 147 college students and 21 adults in evening college extension classes.

Because the process of repeated introspection on the self might be expected to lead to some of the self-insights anticipated in therapy, it was hypothesized that self-concept alterations similar to those reported in successful counseling would be revealed in repeated self-descriptions without psychotherapy. The *Q*-sort methodology was employed to permit comparison with results of recent research projects dealing with changes accompanying therapy which have made use of this technique (2, 3, 4, 7, 8, 9, 10).

The self concept was defined as the individual's phenomenologically unitary constellation of beliefs about and attitudes toward himself, the organization of his self-reflexive affective-cognitive structures,[3] as reflected operationally in his description of himself.

Method

Items for the *Q*-sort self-concept scale were derived from 200 anonymous self-descriptions by university students and urban adults. Self-statements were classed as positive or negative on the basis of ratings by a panel of eight judges. A preliminary set of 180 items was reduced to the final 120, 60 positive and 60 negative, by item analysis of a pilot study with 26 subjects.

Subjects were asked to describe themselves in terms of these items by distributing the self-statements into eleven categories, using a platykurtic seminormalized distribution, shown in Table 1. The same items and the same distribution were employed for description of the self-ideal, defined operationally as the subject's description of the self he would like to be.

This design makes it possible to correlate, item by item, the pattern of self-description

---

[1] Based upon a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree, Vanderbilt University. The writer wishes to express his gratitude to Dr. George E. Copple for his advice and guidance.

A summary of this article was presented at the 1954 meeting of the Southern Society for Philosophy and Psychology.

[2] The writer is now Director of Psychological Services for the Tennessee Department of Mental Health.

[3] Definition formulated by the Vanderbilt-Peabody self-concept research group, directed by Dr. Ted Landsman.

Table 1

Card Distribution in Q-Sort Procedure

Interpretation	Category	No. of cards
Most like me	11	6
	10	9
	9	11
	8	12
	7	14
	6	16
	5	14
	4	12
	3	11
	2	9
Least like me	1	6

for any single individual with his own replications and with the description of the self-ideal. Furthermore, since a score in the range between 1 and 11 points is assigned each item by placing it in one of these 11 categories, a positive self-concept score may be obtained from each sorting by summing the scores assigned the 60 positive items.

In a test-retest reliability study, 120 college students described themselves twice, with the sortings separated by an interval of one week. Mean of the product-moment correlations of individual items for the two sortings for the 120 subjects was .79 ± .02. The reliability coefficient for the positive self-concept scores obtained by summing points assigned the 60 positive items was .95 ± .03.

In testing the hypothesis that changes similar to those reported in psychotherapy would be found for persons engaged in self-introspection by way of repeated self-description, 20 undergraduates and 6 graduate students were used as subjects. Four of these were adults in evening extension classes. The seven subjects in the intensive self-concept group made ten self-sorts in five days, two per day. Fifteen subjects in the self and self-ideal group also made repeated self-descriptive sortings in a five-day period, with two or three self-ideal Q sorts interpolated. In the group of four used for study of changes over a longer period of time, three made ten self-sorts at irregular intervals over periods varying from one to three months in length, while one subject completed 21 sortings in seven different environmental contexts over seven and one-half months.

Results and Discussion

Positiveness of self concept. With the instrument employed in this research, a positive self-concept score in the range 223–497 points may be obtained by placing the 60 positive items in the 11 available categories. If positive and negative items are assigned equal importance in the self-description, the positive score is 360 points with an equal total for the negative items. For the 120 subjects of the reliability group, the mean positive score on the first sorting was 421.25 points, with a range of 262 to 492, and 105 subjects (88%) scoring themselves positively or above the 360-point neutrality total. The mean for the second sorting a week later was 426.78, a mean gain of 5.63 points, statistically significant at the 1% level (Table 2).

The gains were substantially greater for the subjects who made repeated self-sortings in the brief period of five days. The mean increase in positiveness for the seven subjects describing themselves twice daily for five consecutive days was 23 points, a change significant at the 2% level, even for this small *N*. For the 15 subjects interpolating self-ideal descriptions within the self-concept replication series, the mean increment was 17 points, again significant at the 2% level, using in each case the one-tailed test of significance appropriate where the change is in the direction predicted. For the four subjects whose

Table 2

Increase in Positive Self-Concept Score with Repeated Q-sort Self-Description

Group	*N*	Mean		Difference
		First Sorting	Last Sorting	
Reliability group	120	421.25	426.78	5.63*
Intensive self-sorts	7	433.00	455.86	22.86**
Self and self-ideal sorts	15	431.00	447.93	16.93**
Scattered self-sorts	4	436.75	439.50	2.75

* Significant at 1% level.
** Significant at 2% level.

self-sorts were scattered over periods of one to seven and one-half months, with temporal intervals more like those found in research in which the *Q* sort has been used to follow changes taking place in therapy, the mean gain was slightly less than three points (Table 2).

It would appear, then, that self-introspection of the kind induced by repeated *Q*-sort self-description in a short time interval tends to be accompanied by increased positiveness of self-attitude, like that reported for therapy. The change would seem, however, to be of much smaller magnitude than that reported in counseling. The gain in positiveness of self-concept adjustment score noted by Dymond (2) for 25 subjects receiving psychotherapy was approximately 38%, as compared with an 11% increase in positiveness for the subjects of the intensive self-sort group in this research, and 8% for the self and self-ideal group.

Relationship between self concept and self-ideal. Fifteen subjects in the experimental group made repeated self-concept sortings with self-ideal *Q* sorts interpolated within the series and at the conclusion. All of these subjects sorted the items to describe the ideal self at least twice, once following the first self-sort, and once at the conclusion of the series on the fifth day.

The mean correlation between the initial self-description and the first self-ideal sorting was .52, which would seem to indicate a substantial relationship (Table 3). But the range for individual subjects was from .04 to .79, reflecting wide individual differences. Furthermore, the mean positive score for the self-ideal was 493 points, exceeding the mean for the self concept (424) by 69 points, a difference significant at the 1% level. Thus though there seems usually to be a positive relationship between self concept and self-ideal, there is also usually a significant difference.

With replication of descriptions of self and self-ideal over the five-day period, the mean positive correlation between self-concept and ideal self increased from .52 to .66, a gain significant at the 1% level for the 15 non-counseled subjects. This is the trend reported in successful psychotherapy, as illustrated in Rogers' case of Mrs. Oak (9) with the positive relationship between self and self-ideal rising from .21 to .69, and in Hartley's case of Zar (3) with the correlation increasing from .18 at the beginning to .81 at the conclusion of counseling. The increment is much smaller for members of the experimental group than for the two counseling cases, however, with the greatest gain for a subject not receiving counseling being from .43 to .74 in five days. The changes reported in counseling took place, of course, over a period of months. For one subject in the experimental group, Rah, the only one with an initially negative self-concept score of the type often found in persons seeking therapy (330 points), the relationship actually dropped from .04 on the first pair of sortings to − .18 for the last pair, when the positive score had fallen to 315 points.

Table 3

Change in Relationship Between Self Concept and Self-Ideal

Group or individual	Correlation	
	First sorts	Final sorts
Experimental group ($N = 15$)	.52	.66*
Rah	.04	−.18
Mrs. Oak	.21	.69
Zar	.18	.81

* Difference significant at 1% level.

Thus it appears that repeated interacting introspection on the self and the ideal self tends to lead to increased positive relationships like those reported in therapy, though of lesser magnitude. And as in therapy, the description of self tends to be altered more than the self-ideal, though changes took place in both for all subjects.

Consistency of the self concept. Brownfain (1) has reported a significant positive correlation between consistency of self concept, and adjustment as indicated by inventory scores and social relationships. Rogers (9) has noted an increase in consistency of self-description in the case of Mrs. Oak, not only through the therapy period but also in follow-up studies, from an initial correlation of

.50 for the first two self-sorts to .70 for the last pair.

For the experimental subjects in this research project making ten self-sorts in five days, the mean correlation of item scores for the two self-descriptions on the first day was .88. The mean correlation for the last pair of self-concept sortings on the final day was .98. This difference is significant at the 1% level. For the 15 subjects making both self and self-ideal sortings, the increment was from a mean of .79 for the first pair of self-sorts to .88 for the final pair, also significant at the 1% level (Table 4). The change for the four subjects who described themselves at irregular intervals over longer periods of time was from a mean of .80 to .88. Of these 26 subjects, 24 showed steady increases in self-consistency.

Apparently, then, increased consistency of self-description and presumably of self concept is to be anticipated with replication of description of self, whether intensive or spaced over relatively long time intervals, and whether psychotherapy is provided or not.

Summary and Conclusions

To determine whether changes in the self concept similar to those reported in successful psychotherapy may take place with replication of self-description by noncounseled subjects, repeated *Q*-sort descriptions of the self were obtained from 26 persons. In addition, 15 of the subjects interpolated self-ideal sortings within the series of self-sorts. For all conditions employed, the same general trends were found that have been reported for successful counseling. The self concept tends to become more positive and more consistent, and its relationship to the self-ideal becomes more positive. In all cases, however, these alterations were of smaller magnitude than those reported in counseling.

Table 4

Increase in Consistency of Self-Description with Replication

Group or individual	*N*	Mean correlation	
		First sorts	Final sorts
Intensive self-sorts	7	.88	.98*
Self and self-ideal sorts	15	.79	.88*
Scattered self-sorts	4	.80	.88
Mrs. Oak		.50	.70

* Difference significant at 1% level.

The data obtained suggest these conclusions:

1. Intensive self-introspection without therapy appears to be accompanied by increased positiveness of attitudes toward the self, at least for most subjects with initially positive self-attitudes. The increment, however, is usually significantly smaller than that reported in successful counseling cases.

2. Repeated description of the self and self-ideal is usually accompanied by increased positive relationship between the two, even without psychotherapy. The change is smaller than that reported for persons in therapy.

3. Repeated self-description without counseling usually is accompanied by increased consistency of self concept.

4. Self-introspection by self-description without therapy may be accompanied by some of the changes reported in successful counseling, which presumably also involves rather intensive introspection on self. Use of self-descriptive techniques, e.g., the *Q* sort, while therapy is in progress, may then actually facilitate some of the alterations said to be characteristic of successful counseling. These techniques do not appear to produce changes of the magnitude reported for psychotherapy, however.

5. Significant increases in positiveness of self concept, and in positive relationship between the self and self-ideal, may be valid indexes of improvement wrought by therapy, but increased consistency of self concept is achieved so readily by self-description without counseling that it would seem a dubious criterion, especially when self-inventories or *Q* sorts are used in conjunction with therapy for evaluation or other purposes.

Received September 20, 1954.

References

1. Brownfain, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597–606.
2. Dymond, Rosalind F. An adjustment score for Q-sorts. *J. consult. Psychol.*, 1953, **17**, 339–343.

3. Hartley, M. A Q-technique study of changes in the self concept during therapy. Unpublished doctor's dissertation, Univer. of Chicago, 1951.
4. Mowrer, O. H. "*Q*-technique"—description, history, and critique. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research*. New York, Ronald Press, 1953. Pp. 316–375.
5. Raimy, V. C. The self-concept as a factor in counseling and personality organization. Unpublished doctor's dissertation, Ohio State Univer., 1943.
6. Raimy, V. C. Self-reference in counseling interviews. *J. consult. Psychol.*, 1948, 12, 153–163.
7. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
8. Rogers, C. R. Perceptual reorganization in client-centered therapy. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald Press, 1951. Pp. 307–327.
9. Rogers, C. R. Studies in client-centered psychotherapy III: the case of Mrs. Oak—a research analysis. *Psychol. Serv. Center J.*, 1951, 3, 47–165.
10. Stephenson, W. *The study of behavior*. Chicago: Univer. of Chicago Press, 1953.
11. Taylor, D. M. Consistency of the self concept. Unpublished doctor's dissertation, Vanderbilt Univer., 1953.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Self-Evaluations, Group Evaluations, and Objective Measures[1]

Wilse B. Webb

U. S. Naval School of Aviation Medicine [2]

In a previous paper, it was noted that at least three selves could be conceptualized. These were: "(*a*) The objective self, defined in terms of his actual potentialities, capacities, and the like, i.e., the objectively measured self; (*b*) The self concept, the individual's understanding or evaluation of these potentialities and capacities; (*c*) The social self, the operation of the potentialities and capacities in relation to his social environment" (1, p. 307). This previous paper further noted that "It may be assumed that the three levels would be related, but disparity would be possible between all three levels. Unfortunately . . . direct comparison of the self levels from our present data are [*sic*] not possible" (1, p. 307).

The present paper concerns itself with these interrelationships between the three selves described above. Specifically, measures were obtained as estimates of these three selves and the interrelationships are reported and discussed.

Procedure

Four sections of Naval Aviation Cadets with N's of 29, 24, 20, and 22 were the subjects of this experiment. Ratings were obtained from these subjects after they had been members of their section for 4 weeks and 15 weeks. During this time, each member of each section had had constant and close contact with every other member of his section. They had lived, worked, slept, and eaten together almost daily.

The traits on which ratings were obtained at these periods were defined on each rating scale as follows:

Leadership: Could take over and get others to go along with him, could direct a group in accomplishing a goal, could act as a leader.

Social adequacy: Ability to get along with other people; friendly; liked by the members of his group.

Intelligence: A "bright guy"; original; *ability* to meet *new* situations effectively; catches on quickly; can use his head; insightful.

Possibility of success in flight training program.

Possibility of success as a Naval Aviator.

For each trait a rating from one to seven was possible with the extremes labeled as "least" and "most." The cadets were instructed to "normalize" their ratings, that is, they were told to rate the relative standing of their own group rather than rating against a general population. As such, they were told that they should give about as many low ratings as high ratings.

Each member of the section was given a booklet of rating scales which were numbered from 1 through the number in his section. The cadet was also given a roster of his section in which the names had been assigned in a random order to numbers 1 through N. The cadet was told to look at the number on the rating form, to look at the name appearing by this number on the roster, and to proceed to rate that person on the five traits of the scale. He also was told that one of the numbers would appear by his own name. He was to rate himself on these traits and to circle that number on his rating scale.

At the completion of the ratings in the 15-week session, the Otis Quick-Scoring Mental Abilities Test (Form Am) was administered to each group.

---

[1] This study was performed as part of ONR Contract NR 154–098.

[2] Opinions or conclusions contained in this paper are those of the author. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department.

Results

In this paper we have concerned ourselves with only one of the traits of the rating scale, that of intelligence. The correlations between the three measures of intelligence obtained at the 15-week session were:

Group rating and self-rating	.43
Group rating and Otis	.49
Self-rating and Otis	.21

The group rating was the mean of all of the ratings given to an individual by his peers; the self-rating was that rating which the individual assigned to himself; the Otis rating was his score obtained on the Otis test of intelligence.

In addition, it was found that the correlations between the ratings obtained at the 4-week session and at the 15-week session were:

Group rating (4 week) and group rating (15 week)	.76
Self-rating (4 week) and self-rating (15 week)	.19

An analysis of variance estimate of the reliability of the group ratings was .86 at the 4-week session and .89 at the 15-week session.

Discussion

In regard to the group ratings, it is quite clear that the magnitude of these correlations is dependent primarily upon the number of judgments obtained rather than some such factor as the "objectivity" of the rater. It is to be noted that if only one judge was being used, in contrast to approximately 25 judges in each group, correlations of the order of those obtained on the self-ratings would have been expected. These correlations could be estimated by considering each judgment an item, and inverting the Spearman-Brown formula and estimating the reliability of a "test" one–twenty-fifth as long as that used. The resultant reliability would be about .11. Similarly, one could estimate the validity of the group ratings using one–twenty-fifth of the judges by applying the formula:

$$\text{Estimated } r_{\text{criterion}} = \frac{\sqrt{1 - r_{\text{reliability}}}}{n} + r_{\text{reliability}}.$$

In this formula, .04 would be used for the *n*. In this case, the correlation between single ratings of a group member and intelligence would be about .18.

Further, it is possible to surmise that the low reliabilities and perhaps the low validity of the self-ratings were related to this factor of "test length." The demonstrated reliability of personality inventories (which involve self-evaluation but also involve a large number of items) in comparison to the presently obtained reliability of .19 would support such a hypothesis. Certainly, some such hypothesis is preferable to the assumption that the individual's "self concept" is as unstable as the data may suggest.

With these considerations the reported correlations lend themselves to the following interpretations. To the extent to which the "social self" may be conceived of as the combined group judgments of an individual by his peers, this self showed a substantial relationship to an objectively measured self on the dimension of intelligence. Further, to the extent to which a self-statement may be considered related to the "self concept" a remarkably low relationship exists between this self and an objective measure of the trait in question.

We must, however, point to the limitations inherent in these statements. In a sense the measure of the social self is a statistical artifact. The quite discrete individual judgments of the group members are not summed and averaged except perhaps within each individual as he responds to these separate judgments of his peers. Only in voting behavior where opinions are summed do we have a case in which group judgments are actually summed. As for the single rating reflecting the "self concept" we have noted that if we consider this self concept to be a fairly stable state the method here did not show such a stability. Perhaps to get a true picture of the self concept we must obtain successive judgments or multiple judgments related to the trait in question (similar to the personality inventory).

A still further limitation should be noted. We have limited ourselves to peer judgments. These ratings, placed in the matrix of the total social group of an individual composed of not only peers but superiors and inferiors, may reflect quite a different picture. It is quite likely that the actual "selves" of an individual are markedly affected by the varying portions of superiors, peers, and inferiors in his operating environment.

Summary and Conclusions

Four groups of Naval Aviation Cadets who had lived together for 15 weeks rated each other and themselves on a number of traits including intelligence. The Otis test of intelligence was then administered to these individuals. Interrelationships between these measures and ratings obtained after 4 weeks were computed. It was found that the averaged group ratings were reliable and correlated substantially with the objective measures of intelligence. The self-ratings appeared to be highly unreliable and exhibited only a limited relation to intelligence. It was noted that the reliability and possibly the validity of these relationships were considerably dependent upon the number of judgments obtained. That is, the group ratings were an average of individual ratings which were no more reliable than the self-rating. It was further noted that the group ratings were a statistical phenomenon and not an actual thing. Further, it was noted that our findings were within a peer context and the effect of presence of superiors and inferiors on these ratings was not known. In spite of these limitations, and there are probably others, two empirical facts could be stated:

1. A group of peers can fairly well "classify" its members on intelligence. The way a person performs in relation to his peers results in a reliable and valid "social self" in regard to intelligence if this social self can be defined as the summation of the group's attitude toward the person.

2. An individual's single statement about his intelligence is likely to have little relationship to his measured intelligence at that time in a group of peers. Stated otherwise, if a self-statement is related to the "self concept" of intelligence this self concept is likely to be highly unrealistic in relation to objective measures.

Received September 21, 1954.

Reference

1. Webb, W. B. Self-evaluation compared with group evaluations. *J. consult. Psychol.*, 1952, **16**, 305–307.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

MMPI Validity Scores as a Function of Increasing Levels of Anxiety

Joseph D. Matarazzo
Washington University School of Medicine

In an earlier study it was noted, "Our work with the Taylor (anxiety) scale has led us to wonder if the individuals in the low anxiety group (our group I) are not *in fact* equal in anxiety level to the *S*s in the high anxiety groups but are denying (by their answers to the Taylor) this anxiety. That is, they might be equal in level of anxiety to the other groups but are repressing or suppressing this fact" (9). However, the finding that low anxious and high anxious *S*s earned different learning scores (trials) in a maze learning task suggested that the two groups might truly be different and that denial was not operating in the low group.

The present study represents an attempt to investigate this question further. The Taylor Scale of Manifest Anxiety (13) is a questionnaire-type inventory, derived from the MMPI, containing 50 critical items which constitute the anxiety scale and 175 filler items. Fortunately, for our purposes, these 175 filler items contain the three validity scales, *L*, *F*, and *K*, of the MMPI (3), which are described in more detail below. The aim of the present study was to determine the strength of the relationship between scores on these three validity scales and level of anxiety as judged from the individual's score on the Taylor scale.

Procedure

The *S*s used in the first part of the study were 119 male medical freshmen and sophomores who had served in an earlier study (11). On the basis of their Taylor raw scores the 119 medical students were assigned to five groups on a scale of increasing anxiety level. Individuals with raw scores of 1–5 were given an anxiety rating of 1, while those with raw scores of 6–10, 11–15, etc. were given ratings of 2, 3, 4, and 5, respectively. The number of *S*s in each of these five groups from low to high was 18, 41, 34, 15, and 11, respectively. The findings with this group were cross validated the following year with a group of 31 female occupational therapy (O.T.) juniors. The Taylor scale was group administered in both samples.

Results

In Fig. 1 is shown the relationship for the medical students between the *L* score and anxiety level as measured by the Taylor scale. The strength of this relationship is a moderately negative one; the value of the Pearson r is $-.32$, a value significant at the .01 level of confidence. Values of eta and

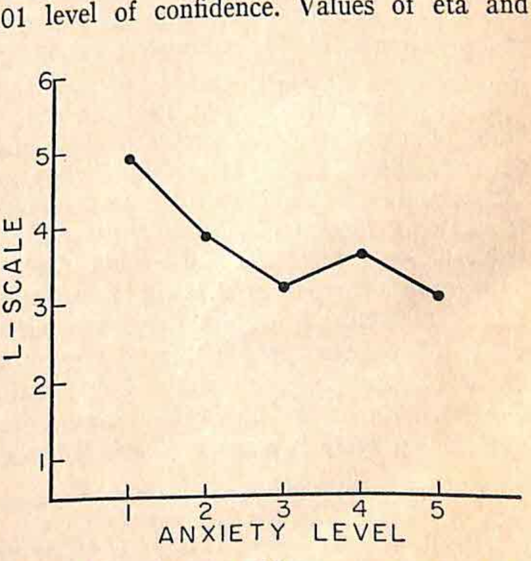

Fig. 1. Mean Lie-scale (L) score for each of the five groups ranging from low anxiety (group 1) to high anxiety (group 5).

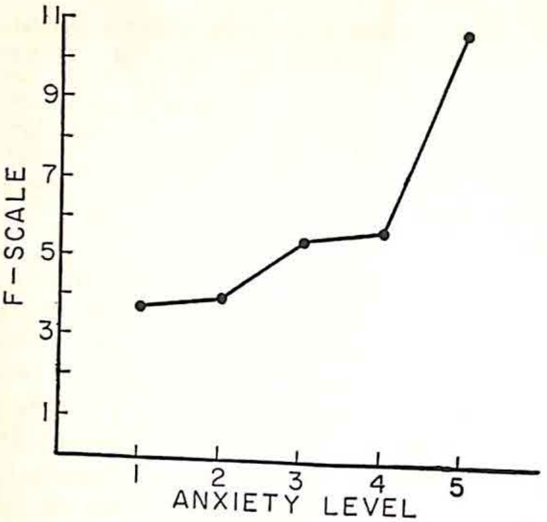

Fig. 2. Mean *F*-scale score for each of the five groups ranging from low anxiety (group 1) to high anxiety (group 5).

epsilon, computed from the analysis of variance data, yield similar results, .36 and .32, respectively. Subjects in group 1 earn a Lie score mean of approximately 5.00 while those in group 5 earn a mean of 3.00. Thus there is a moderate tendency for those *S*s scoring low on the Taylor scale to be less candid (as inferred from their score on the *L* scale) than individuals in the higher anxiety ranges. This relationship was even more striking with the O.T. sample. The value of r was found to be $-.52$ with this female sample. The means ranged from 5.71 in group 1 to 2.40 in group 5.

The relationship between score on the *F* scale[1] and anxiety level for the medical students is shown in Fig. 2. The strength of this relationship, significant at the .01 level, is greater than that of *L*, as seen from the r value of .46 (eta of .50 and epsilon of .46). The means vary from 3.72 in group 1 to 10.73 in group 5. Again this relationship tends to be stronger in the O.T. sample (r of .62 and approximately similar mean values). From these findings it can be concluded that there is a moderate tendency for whatever it is that is measured by the *F* scale to be correlated with the Taylor index of anxiety level.

In Fig. 3 is shown the relationship, again for the medical students, between anxiety level and score on the *K* scale. The magnitude of the correlation, r of $-.71$ (eta of .69 and epsilon of .71), indicates a very high relationship (.01 level) between these two measures. The mean values of *K* vary from approximately 20 in group 1 to 11 in group 5. The correlation is again stronger in the O.T. sample: r of $-.84$, and means ranging from 22 to 10. This very high relationship between the *K*-scale index and score on the Taylor scale would indicate that one's anxiety level as inferred by his Taylor score reflects his "test-taking attitude" as well as whatever else it is the Taylor scale is sampling.

Discussion

The correlations obtained between the Taylor questionnaire and each of the three validity scales raise a number of interesting questions. The interpretation of these correlations would be less complicated if the three validity indices represented "independent" measures and were not themselves composed of items from the 550-item MMPI. We are faced here with the problem of interpreting correlations between two scales (the Taylor and each of the three validity scales) when

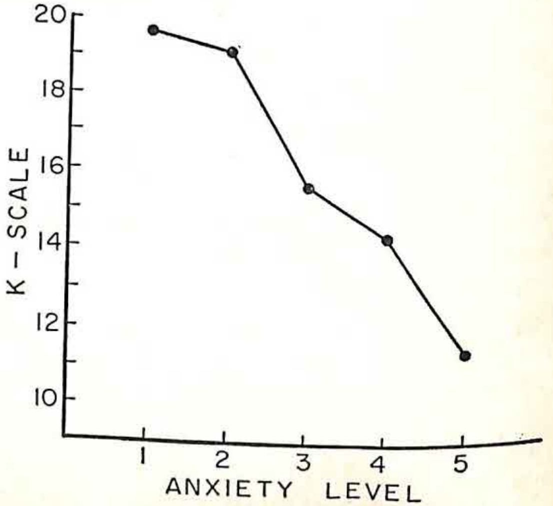

Fig. 3. Mean *K*-scale score for each of the five groups ranging from low anxiety (group 1) to high anxiety (group 5).

---

[1] Two of the 64 items of the *F* scale were apparently not included in the Taylor scale but this is considered relatively negligible for our purposes. Also there is an overlap of three items between the Taylor and the *F* scale. This indicates that the obtained r between these two measures is slightly greater than if this overlap did not exist.

both of these scales are composed of "personality-tapping" questionnaire-type items from a single population of items. A way out of this dilemma is to pay less attention to what it is these scales are purported to measure and to think in terms of the operational procedures involved in their original development and standardization.

The Taylor scale consists of 50 items, chosen out of a total of 200 MMPI items, upon which 5 clinicians showed 80 per cent agreement that these were indicative of manifest anxiety as the latter was described by Cameron. The literature of the past four years indicates that scores on this scale are associated with differences in conditioning and learning ability as demonstrated in numerous studies (8), intelligence as measured by a timed test (10), and also that these scores can differentiate, fairly efficiently, medical from psychiatric patients (12). The Lie scale (*L*) is composed of 15 items of the following sort: I would rather win than lose in a game; at times I feel like swearing; I do not always tell the truth; I gossip at times; etc. The authors of the MMPI found empirically that a large, representative sample of "normals" will usually answer "yes" to these and the other similar items making up the scale. The *F* scale was similarly developed and consists of 64 items which are answered in the infrequent direction less than 10 per cent of the time by these same "normals." A subject's score on the *F* scale (3) has been found empirically to be a function of carelessness in taking the test, inability to comprehend the items, errors in recording on the answer sheet, the presence of psychiatric illness (1, 6), and extreme individuality (in the sense that it picks up those few people who are highly individual and independent and who honestly score those items which are infrequently scored by "normals"). The *K* scale, a 30-item scale purported to assess defensiveness or lack of it in a subject's "test-taking attitude," was empirically derived by studying the item response frequencies of certain diagnosed abnormals whose MMPI pattern yielded essentially normal profiles. It was assumed by the MMPI authors that the occurrence of a normal profile in a known psychiatric patient was suggestive of a defensive attitude in this patient's response behavior. The response frequencies of these "normal appearing" patients were contrasted with those from an unselected sample of people in general ("true normals"). The items which differentiated these two groups were then scored so that a high *K* score would be found among known abnormals with normal curves, whereas a low score would be found in clinical normals having deviant curves. The authors then identify the trait sampled by the *K* scale by suggesting that, in this operational sense, it can be said that a high score is indicative of a defensive attitude, and a low *K* score suggests unusual frankness or self-criticalness (7).

The authors of the MMPI consider all three of these scales in two ways: (*a*) as validating scales, the scores of which will help the clinician determine for any subject whether or not his responses were unduly influenced by his test-taking attitude, and (*b*) as an index of certain aspects of his personality, i.e., if he rejects by his test responses all those items that imply personal faults or personality difficulty then this very rejection (reflected in the validity-scale scores) tells us something about the kind of person he is (3, p. 23).

It is in regard to the second interpretation of the validity scales that the results of the present study are pertinent. The correlations obtained between the Taylor scale and each of these three measures serve the purpose of bringing together two separate streams of research—studies growing out of hypotheses based on learning theory and clinical experience (Taylor scale) are thus related to the many personality theory studies which have grown out of the parent instrument, the MMPI.[2] If one brings together the results of these two independent streams of research, it would appear that *S*s who score low on the Taylor scale and who have been shown to be superior in learning ability (to *S*s whose anxiety level is higher) on some tasks, and to be inferior on still others, and who earn higher scores on the ACE test of intelligence, etc., are also individuals who are not as frank and are probably more defensive in their test-taking behavior. Individuals who score in the middle range or at the high end on the Taylor scale can be similarly understood as progressively less defensive. Thus, in the sense of the second usage of the validity scales, the correlations obtained in this study enable us to know more about the characteristics of an individual subject than would be possible from knowledge of his Taylor score or his scores on the validity scales alone. The implications for the Taylor scale, and studies based on it, of these findings (especially the lack of "frankness" in the low "anxious" groups) will become clear only after further research.

These data would suggest, however, that the superiority of the high anxious group (Taylor scale) in numerous conditioning studies reported to date is as much a function of their "lack of defensiveness" in test-taking attitude as it is a function of their anxiety level. The performance of the middle and low anxious groups can be similarly interpreted. In this connection it is of interest that Baron [3] has found a "striking curvilinear relationship" between the MMPI *K* score and verbal learning. The high correlation found between *K* and anxiety scores would thus indicate that Baron's results confirm our earlier finding of a significant curvilinear relationship between anxiety score and stylus maze learning (9). The significant correlations found in the present study raise some other interesting questions: Is the Taylor scale a measure of "anxiety," or is it a measure of "defensiveness," or lack of "candidness"? Are the *K* and *L* and *F* scales also a measure of anxiety? Or is it possible that all four scales are tapping a common factor, a factor we might tentatively call "level of maladjustment." Further research will be needed to answer these questions, although it is of interest that the Taylor scale (as well as the validity scales) can differentiate, reasonably efficiently, a psychiatric from a nonpsychiatric patient sample (12).

---

[2] It is of interest to recall that the Taylor scale represents items which *clinicians judged* indicative of manifest anxiety, while the three validity scales are composed of items which *empirically have separated groups* (patients from "normals," malingerers from nonmalingerers, etc.) and in which clinical judgment played no a priori role.

The findings of the present study are also of interest for the MMPI user. The grand mean values for all 119 medical student *S*s of this study (independent of Taylor anxiety level) on the *L*, *F*, and *K* scales were 3.71, 5.17, and 16.87, respectively. These are mean values which are clearly in the established limits considered "normal," since raw scores of 7, 17, and 23, respectively, are necessary to earn a *T* score of 70 or above. The curves shown in Fig. 1, 2, and 3 therefore represent a finer analysis of validity scores (heretofore considered "normal") along a dimension of anxiety. Thus the correlations permit a refinement of the "meaning" of the validity scores.

The question can be raised whether the demonstrated relationships between the items making up the validity scales and those making up the Taylor scale are exact ones since the three validity scales, imbedded in the 225-item Taylor questionnaire, might be responded to differently than if the full 550-item MMPI were administered. While this question must be left open, as it was not specifically investigated in this study, it is felt that the effect on the demonstrated correlations, if any, would not be serious.

The results of several recent publications are also of interest in relation to our findings. Heineman (4), working in the Iowa laboratory where the Taylor scale was developed, set out specifically to develop a forced-choice form of the Taylor inventory in order to reduce the effects of the desire of *S*s to appear in a favorable light and so earn by their responses lower Taylor anxiety scores. His study was prompted by several findings at Iowa, one of which was a heretofore unpublished correlation of − .74 between the anxiety scale and the *K* scale of the MMPI. The latter is comparable to the correlations found in the present study. By use of his forced-choice technique, Heineman was able to reduce the correlation between Taylor score and *K* to approximately − .55 when he scored his new scale in one way and − .40 when he scored it by a second key. In his review of the Taylor scale, Mandler (8) reports two studies from the Harvard laboratory. In the first place, Ericksen found a correlation of .72 between psychasthenia-hysteria (MMPI hysteria score minus psychasthenia score) and the Taylor, with high anxiety scores related with psychasthenia. In the second, Ericksen and Davids report a rank-order correlation of .92 between the MMPI psychasthenia scale and the Taylor score, and a rank-order r of − .89 with the hysteria-psychasthenia measure. These authors also report an interesting correlation of − .90 between the Taylor scale and rankings on the trait of "optimism-pessimism." They suggest that on the basis of clinical appraisal, high and low scoring *S*s on the Taylor scale also show differential defense mechanisms and self-attitudes. This suggestion is clearly similar to the finding which prompted our own study. Hovey, in a clinical study using psychiatric patients, found that his 60 *S*s who were diagnosed "anxiety reaction" earned the lowest scores on the *K* scale and the highest on the *F* scale (5). These patients thus were similar to the "normal" *S*s in our high anxiety group, group 5, who also scored lowest on the *K* scale and highest on the *F* scale. Deese, Lazarus, and Keenan (2) report a correlation of .81 (reduced to .40 when overlapping items were excluded) between the Taylor and the MMPI psychasthenia scale. All of these studies suggest that there are many interesting relationships between these MMPI clinical scales and anxiety as reflected in the Taylor score (or diagnosed clinically, as in the study by Hovey).

One further point is worth noting. The scores of research papers utilizing the Taylor scale published since this crude index of anxiety was introduced in 1951 are an indication of the eagerness with which psychological science awaited a reliable, and, on the face, valid index of anxiety. The high correlations (− .71 and − .84) obtained in this study between the Taylor and *K* scale would indicate that the latter may be used as an *alternate* form of the Taylor (with appropriate consideration of the negative direction of the correlation) in research requiring two forms of the scale. These correlations between *K* and the Taylor scale are not much lower than the two reliability coefficients of the latter (.81 and .89) reported by Taylor (13).

---

[3] Baron, M. R. Personal communication, October 1953. Professor Baron also found a high r between *K* and anxiety.

The trend for the all-female O.T. sample to show consistently *higher* correlations between the Taylor scale and each of the three validity scales, as compared to the all-*male* (possibly more intelligent) medical student sample, is difficult to interpret. Further research may offer an hypothesis to explain this observation.

Summary

The Taylor anxiety scores of 119 male medical students were correlated with their scores on the three validity scales (*L*, *F*, and *K*) of the MMPI. The obtained correlations were − .32, .46, and − .71, respectively. These values were all significant at the .01 level of confidence. These findings were cross validated on a sample of 31 female occupational therapy students. There was a consistent trend for all correlations to be higher in this second sample. It was suggested that these correlations serve the purpose of bringing together two heretofore separate streams of research—Taylor scale studies growing out of hypotheses based on learning theory and clinical experience, on the one hand, and personality theory studies which have grown out of the parent instrument, the MMPI, on the other. Questions regarding the interpretation of what each of these scales "taps" were raised.

Received October 11, 1954.

References

1. Brožek, J., & Schiele, B. C. Clinical significance of the MMPI F-scale evaluated in experimental neurosis. *Amer. J. Psychiat.*, 1948, **105**, 259–266.
2. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.
3. Hathaway, S. R., & McKinley, J. C. *Minnesota Multiphasic Personality Inventory. Revised manual.* New York: Psychological Corp., 1951.
4. Heineman, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447–454.
5. Hovey, H. B. Somatization and other neurotic reaction and MMPI profiles. *J. clin. Psychol.*, 1949, **5**, 153–157.
6. Kazan, A. T., & Sheinberg, I. M. Clinical note on the significance of the validity score (*F*) in the MMPI. *Amer. J. Psychiat.*, 1945, **102**, 181–183.
7. McKinley, J. C., Hathaway, S. R., & Meehl, P. E. The MMPI: VI. The *K* scale. *J. consult. Psychol.*, 1948, **12**, 20–31.
8. Mandler, G. Anxiety and performance: empirical and theoretical aspects of anxiety scales. Unpublished manuscript, Laboratory of Social Relations, Harvard University.
9. Matarazzo, J. D., Ulett, G. A., & Saslow, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, in press.
10. Matarazzo, J. D., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, **18**, 201–205.
11. Matarazzo, J. D., & Phillips, Jeanne S. Digit symbol performance as a function of increasing levels of anxiety. *J. consult. Psychol.*, 1955, **19**, 131–134.
12. Matarazzo, J. D., Guze, S. B., & Matarazzo, Ruth G. An approach to the validity of the the Taylor Anxiety Scale: scores of medical and psychiatric patients. *J. abnorm. soc. Psychol.*, in press.
13. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

The Relationship of Manifest Anxiety to Wechsler-Bellevue Subtest Performance[1]

Ruth G. Matarazzo
Washington University School of Medicine

Since the Taylor scale is an objective and apparently valid measure of anxiety, it seemed that here was a way, free of the usual hazards of subjectivity and bias, for selecting groups on which to test some common clinical assumptions about the subtest pattern of the Wechsler-Bellevue to be expected as a result of anxiety.

Wechsler does not report the characteristics or size of his neurotic group but, apparently on the basis of clinical impression, describes it as tending to be particularly low on the Digit Span and Object Assembly subtests. In their two volumes Rapaport, Gill, and Schafer reported a similar pattern based, however, on an Anxiety and Depression group consisting of only ten patients.

The subjects for the present study were 80 medical and psychiatric hospitalized, white, male patients between the ages of 18 and 37. The subjects were administered the Wechsler-Bellevue and asked individually to fill out the Taylor scale. Raw scores on the Taylor scale were coded into five anxiety groups, Taylor scores from 1 to 8 receiving an anxiety rating of one, those from 9 to 16 a rating of two, and so on. The number of subjects within each group from 1 to 5 were 7, 17, 16, 21, and 19, respectively. F tests indicated that the groups were equated for age, education, and IQ, the mean age being 29, mean education 10th grade, and mean IQ 105.

By means of the F test the performance of the five groups was compared on each of the eleven subtests of the Wechsler-Bellevue and no relationship was found between anxiety level and subtest performance. The use of a correlational technique (epsilon) confirmed this lack of relationship; no one of the eleven epsilon values reached significance.

Since the group means might be obscuring true differential patterning effects or "scatter" among the various subtests for *individual* patients, two measures of individual scatter analysis were used: (*a*) range; and (*b*) deviation from the mean (a count was made of the number of subtests which differed by three or more weighted score units from the individual's mean weighted subtest score). Neither of these two measures bore a relationship to anxiety level.

Comparing the upper and lower 20% of the anxiety continuum, as has been done in most studies with the Taylor scale, did not change the negative results. Eliminating subjects with IQ less than 100 also produced no change.

Assuming that the Taylor scale measures the same thing which clinicians mean when they speak of anxiety, the present results must lead us to carefully question the validity of the diagnostic signs described by Wechsler, Rapaport, and others. The present results are perhaps surprising in view of the long and widely held belief among clinicians that digit span especially, and perhaps object assembly to a less extent, are vulnerable to anxiety.

Brief Report
Received December 21, 1954.

---

[1] An extended report of this study may be obtained without charge from Ruth G. Matarazzo, Dept. of Neuropsychiatry, Washington University School of Medicine, St. Louis, Mo., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4523 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

The Relation of Manifest Anxiety to Association Productivity and Intellectual Attainment[1]

Anthony Davids and Charles W. Eriksen
Harvard University

Within Hull's theoretical system all habit tendencies activated by a given stimulus are considered to be multiplied by the total drive then operating. This proposition has been subjected to a number of recent experimental tests. A popular technique has been to employ a personality scale of manifest anxiety (8) to provide a measure of total drive strength. Performance of *S*s selected on the basis of high or low scores on this scale has been compared on such measures as eyelid conditioning (2, 5, 6, 7), verbal learning (3, 4, 9), and various other more complex tasks (1, 10, 11). In general, the results have tended to support predictions derived from Hullian theory.

This theoretical formulation holds two somewhat related predictions for performance measures. The higher the total drive state, the more likely is the occurrence of the strongest response tendency, but, also, with high drive states, the greater is the total number of response tendencies that are above threshold. The studies on eyelid conditioning have rather directly verified the effect of drive level upon a single dominant response, and the verbal learning studies have indirectly confirmed the related prediction that with high drive a greater number of response tendencies would be above threshold. In these latter studies the support has been indirect since it has come from the deduction that high drive leads to a greater number of competing responses which, in turn, differentially interfere with learning depending upon the degree of learning (relative strength of the correct response) that has occurred.

It was our first purpose in the present experiment to provide a more direct test of the deduction that a high drive level would lead to a greater number of responses being suprathreshold. From this deduction, it would be expected that scores on the manifest anxiety scale would be directly related to the number of associations given in response to stimulus words on a chained word association test.

A second purpose in this study was to investigate relations between scores on the anxiety scale and two measures of intellectual ability. Although it is common practice to select subjects from a college population on the basis of their anxiety scores and to compare the learning rate of "anxious" and "nonanxious" groups, the question of whether this selection procedure introduces ability or intellectual differences is as yet unanswered.[2] In the attempt to secure evidence bearing upon this question, we correlated results on the anxiety scale with performance on a battery of college aptitude tests and with academic achievement.

Method

Subjects. Forty male undergraduates participated in this study. They were selected from among a larger number of students who had volunteered to serve in an extensive series of psychological studies. Their selection was on the basis that the group include as much diversification as possible, within the limits

---

[1] This investigation was facilitated by research grants from the Harvard Laboratory of Social Relations, the Rockefeller Foundation, and the National Institute of Mental Health (Grant M-700), Public Health Service.

[2] It should not be overlooked that an unambiguous answer to this question may be difficult, or impossible, to attain since anxiety could conceivably affect performance on "ability" or "intelligence" tests.

of a college population, in regard to socioeconomic status, interests, abilities, and extracurricular activities.

Word association test. This test consisted of a list of 100 nouns, selected as representative of a wide variety of attitudes, beliefs, and personality traits. Ten of the words were designed to relate to anxiety or fear. The words were presented to *S*s by means of a tape recorder. The *S*s were supplied with numbered data sheets and were instructed to write down as many associations as possible to each of the 100 stimulus words. The words were played one at a time from the tape recorder, and 20 seconds were allowed after each word for *S*s' chained associations.

Administration of the manifest anxiety scale. In a session following the word association test, *S*s were individually administered the anxiety scale. The items selected by Taylor (8) were interspersed among the items from the hysteria, psychasthenia, lie, and *K* scales of the MMPI. Scores on the anxiety scale for the 40 *S*s in our sample ranged from 2 to 41 with a median of 17.

Measures of scholastic ability. Transcripts of *S*s' academic records were secured from the registrar's office, and they were rank ordered in terms of their grade-point average. Letter grades, with plusses and minuses taken into account, were transformed to numerical equivalents employing a scoring scheme that ranged from 1 to 12, with "A" equal to 1 and "E" equal to 12. Grade-point averages for the 40 *S*s ranged from 1 (A) to 8 (C −), with a median of 5 (B −), indicating considerable diversity in academic achievement. A separate rank order was determined on the basis of *S*s' performance on the college entrance examinations. Here again pronounced variability was evident, with scores ranging from 474 to 730 and a median of 611.

Results and Discussion

Table 1 contains correlations between scores on the anxiety scale and several measures of performance on the word association test. The coefficients in the first row are based upon the entire series of 100 stimulus words, while the coefficients in the second row are based only upon responses to the 90 nonanxiety stimulus words. In the first column of this table, it can be seen that scores on the anxiety scale are directly related to productivity of associations. Whether or not the 10 anxiety stimulus words are included in the computations, high anxiety scores are correlated positively with a high number of associations. These relationships are significant beyond the .01 level.

Table 1

Product-Moment Correlations Between Manifest Anxiety Scores and Word Association Test Performance ($N = 40$)

Measure	Number of associations	Number of anxiety responses	Percentage of anxiety responses
100 Stimulus words	.45**	.54**	.34*
90 Nonanxiety stimulus words	.40**	.42**	.35*

* Significant beyond the .05 level.
** Significant beyond the .01 level.

The magnitude of the differences in number of associations can be seen when a comparison is made between the means obtained by the 20 highest *S*s and 20 lowest *S*s on the anxiety scale. The mean number of associations given by the high group was 530 and the mean number given by the low group was 454. Comparison of the difference between these means results in a t of 3.21 which is significant beyond the .001 level for a one-tailed test. These findings are in keeping with the expectation that high drive level, as measured by the anxiety scale, increases the number of responses that are above threshold.

While our major concern was whether the anxiety scale would relate to productivity of associations, the present data afford an opportunity to examine the relation between anxiety measured by the Taylor scale and anxiety assessed by the word association technique. Since the Taylor scale contains a large number of items concerning subjective feelings of anxiety, it seems plausible to expect that *S*s scoring high on this inventory would evidence a relatively large amount of verbal content indicative of anxiety and anxiety ideation. To explore this possibility, we computed the number of associations given by each *S* that referred to anxiety ideation.[3] The coefficients in the second column of Table 1 show the degree of correlation between anxiety scores and number of anxiety responses elicited from the different *S*s. Again relations significant beyond the .01 level are obtained whether or not the 10 anxiety stimulus words are included in the computations. The *S*s with high manifest anxiety scores tend to give more anxiety associations than do the low-scoring *S*s.

This latter finding holds true even when the number of anxiety responses is corrected for differences in total productivity. In column 3 of Table 1 the number of anxiety responses has been expressed as a percentage of the individual's total number of responses, and the obtained correlations remain above the .05 level of statistical significance. As an additional control for the influence of total productivity on the relation between the manifest anxiety scale and anxiety associations, partial correlations were computed between these two measures of anxiety, with the total number of associations held constant. In response to the entire word association test and the 90 non-anxiety stimulus words, the partial correlation coefficients are significant at the .05 level (.36 and .31), providing clear-cut evidence of a positive relation between these two diverse approaches to the measurement of anxiety.

The second purpose in this study was to determine whether the relations that have been reported between the Taylor anxiety scale and various performance measures might not have been due in part to a correlation of this scale with intelligence. With measures of scholastic aptitude and achievement as indices of intellectual capacity, college entrance examination scores and grade-point averages were correlated with scores on the Taylor scale. It was found that anxiety scores correlated − .02 with entrance examinations and .17 with grade-point averages. Thus, in neither case did the coefficient approach statistical significance. These two measures of intellectual ability were also correlated with word association test productivity and, again, there was no evidence of significant relationships.

This failure to obtain relationships even approaching significance between results on the anxiety scale and measures of intellectual ability strongly suggests that scores on the scale are largely independent of intelligence. Of course, it must be realized that these correlations were obtained on the basis of results secured from an intellectually homogeneous population. And it should not be overlooked that if the sample contained a representative proportion of the general population a significant relation between anxiety score and intelligence might be demonstrated. But, in this regard, it is important to note that most of the findings reported with the anxiety scale have been obtained with college students similar to those employed in the present study. Therefore, it would seem safe to conclude that the demonstrated differences in performance between *S*s who score high and low on this scale of manifest anxiety are probably not attributable to differences in intellectual ability.

Summary

In the present study, scores on the Taylor scale of manifest anxiety were correlated with performance on a 100-word chained association test. Supporting the prediction based on the supposition that anxiety measures drive, significant positive correlations were found between anxiety scores and productivity of associations. It was also found that *S*s scoring high on the anxiety inventory tended to give relatively more associations containing anxiety ideation than did low-scoring *S*s. Correlation of anxiety scores and association productivity with grade-point averages and performance on college entrance examinations indicated that, in the present sample, both anxiety and productivity were independent of these measures of intelligence.

Received March 21, 1955.
Early Publication.

References

1. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
2. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.

---

[3] The percentage of agreement between two independent raters' scoring of anxiety responses for 20 *S*s was 91.

3. Lucas, J. D. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, **65**, 54–66.
4. Montague, E. K. Role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
5. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
6. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
7. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 91–92.
8. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
9. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–66.
10. Wesley, Elizabeth L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–134.
11. Westrope, Martha. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.*, 1953, **48**, 515–524.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

A Failure to Replicate the Finding of a Negative Correlation Between Manifest Anxiety and ACE Scores

R. E. Schulz and Allen D. Calvin
Michigan State College

In a recent study which appeared in this journal Matarazzo, Ulett, Guze, and Saslow report (1, p. 203), "Therefore it can be concluded that anxiety appears to be moderately, although significantly, correlated with total ACE score and that the higher the anxiety level the lower, on the average, is the obtained score on this test." They obtained a Pearson r of $-.25$ which with an N of 101 was significant at the 1% level of confidence. This finding of course is of considerable importance because of the large number of studies which have recently appeared in the literature in which the Taylor Manifest Anxiety Scale (A scale) was used to separate anxious from nonanxious *S*s. If the A scale is related to the ACE then the differences on learning problems, perceptual tasks, etc. obtained between anxious and nonanxious *S*s chosen on the basis of the A scale may be merely due to intelligence[1] rather than anxiety. With these facts in mind we decided that an attempt should be made to substantiate their findings.

Our sample originally consisted of 99 Michigan State College students, but one student was not included in our analysis because of his extremely low ACE[2] score of 27. Following the procedure utilized by Matarazzo *et al.*, we divided our *S*s into four groups on the basis of their *A*-scale scores. Group I consisted of those *S*s with *A*-scale scores from 1–8; Group II, 9–16; Group III, 17–23, and Group IV, those *S*s with *A*-scale scores of 24 and higher. A Pearson r was computed, and a correlation of .02 was obtained. This practically zero relationship fails to confirm the findings of Matarazzo *et al.* The present authors felt that a more precise estimate might be obtained if the actual *A*-scale scores for each *S* were utilized instead of grouping them into four categories as Matarazzo *et al.* did. We therefore computed an r using individual *A*-scale scores, but again we obtained a zero correlation.

In order to test for rectilinearity an eta was computed, and a value of .10 was obtained which is also not significantly different from zero. A χ^2 test of goodness of fit was made, and the resulting χ^2 fell far short of significance indicating a rectilinear relationship.[3] This supports the findings of Matarazzo *et al.* who also report a rectilinear relationship.

Why did we find a low positive relationship which does not differ significantly from zero while Matarazzo *et al.* report that they found a significant negative relationship? One possibility is that our *S*s had a different distribution of Taylor scores than theirs. They give the following distribution of anxiety scores: Group I had 13 *S*s, Group II, 47 *S*s, Group III, 23 *S*s, and Group IV, 18 *S*s. Our

---

[1] Matarazzo *et al.* failed to find a significant relationship between the Taylor and the CVS. They attribute this lack of relationship to the fact that the CVS is not a timed intelligence test while the ACE is.

[2] The raw ACE score is not given to a student at Michigan State College; instead, a rank is assigned. Therefore, in making our analysis we used the midpoint of the raw scores of each rank except for the highest rank where, of course, only the lower limit was used.

[3] Since the r and eta we obtained are both insignificant, our χ^2 test, of course, is not too meaningful.

*S*s scored as follows: Group I, 24 *S*s, Group II, 44 *S*s, Group III, 22 *S*s, and Group IV, 8 *S*s. A χ^2 was computed, and since it fell short of significance ($p > .05$) the null hypothesis tends to be confirmed. Another possibility is that the difference between their findings and ours could be attributed to variations in selection procedures which might result in a different type of student body at the respective institutions. However, until more evidence appears to support the results of Matarazzo *et al.* it would seem that a valid relationship between intelligence and scores on the Taylor Manifest Anxiety Scale has yet to be established.

Received September 9, 1954.

Reference

1. Matarazzo, J. D., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, **18**, 201–205.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: Addendum

Betsy Worth Estes
University of Kentucky

The present study is a continuation of "Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: An Exploratory Study" (1).

In the previous study there were two groups of subjects: one aged 7 years, plus or minus 2 months, in the second grade ($N = 40$); the other aged 10 years, plus or minus 2 months, in the fifth grade ($N = 40$). Each group was further divided by socioeconomic status into two groups, one of upper economic and one of lower economic status. Each group had an equal number of boys and girls. Mean scores on the WISC did not vary with socioeconomic level in the fifth grade but did vary in the second grade. In the second graders, the upper group exceeded the scores of the lower group at the 1 per cent level. Several possible explanations for this decrease in the effects of socioeconomic status with increase in age and grade were offered.

In 1953–1954, the previous second graders were in the fifth grade; therefore, it was decided to retest them.

Subjects and Procedure

Some of the subjects were unavailable as they had moved out of the state. In the 1950–1951 study there were 20 in the upper group and 20 in the lower. In the 1953–1954 study there were 18 in the upper and 14 in the lower. These latter subjects who were available for the present study had shown a significant difference in the previous study between the upper and lower groups. The WISC was administered to each child.

Results

There was an increase in mean IQ score for each economic group, but the difference between the groups was not statistically significant, as shown in Table 1. The scaled scores of the WISC also showed no significant difference in the 1953–1954 study.

Table 1

WISC Intelligence Quotients of Two Socioeconomic Groups

Socio-economic group	1950–1951			1953–1954		
	Mean	*SD*	*t*	Mean	*SD*	*t*
Upper	114.50	9.49	3.02**	119.44	14.59	1.76
Lower	104.21	9.05		111.93	8.69	

** Significant at less than the 1% level.

Discussion

The same result was obtained in the later study as in the first one: a decrease in the effects of socioeconomic status with increase in age and grade. The evidence is strengthened by the demonstration that the same children who differed in the second grade did not differ in the fifth grade.

It must not be overlooked that the means of the upper group on both IQ's and scaled scores are higher than the means of the lower group. However, by the fifth grade, they did not differ significantly for these subjects.

The possible explanations offered in the first study may still be held: selective factors in promotion, the public schools, movies, Sunday schools, etc. exerting an increasing effect with advancing age and thereby narrowing the differences between socioeconomic levels, or factors within the WISC itself which operate to reduce socioeconomic effects with age.

Summary

Two groups of subjects, 18 in the upper socioeconomic group, 14 in the lower group, were retested on the WISC. The significant difference between the two groups which was found three years earlier when they were in the second grade no longer existed. With increase in age and grade, there was a decrease in difference between groups as measured by mean IQ and mean scaled scores on the WISC.

Received October 11, 1954.

Reference

1. Estes, Betsy W. Influence of socioeconomic status on Wechsler Intelligence Scale for Children: an exploratory study. *J. consult. Psychol.*, 1953, **17**, 58–62.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

Zaslow's Test of Concept Formation: Reliability and Validity[1]

Sol Kugelmass
Psychiatric Institute, Columbia University

and Marvin R. Foundeur
Teachers College, Columbia University [2]

This study is an outgrowth of research on the development and selection of measures for a battery designed to establish a psychological prognosis for schizophrenic patients. A preliminary report by Zaslow (2), in which he described a sorting test used to investigate disturbances in concept formation in a group of schizophrenic subjects, was felt to be sufficiently encouraging to warrant further investigation. Although a recent article by Stacey and Cantor (1) supplied data on subnormal adolescents, to date no study has reported on the reliability of the procedure. To this end, an attempt was made in the present study to evaluate the test's reliability with two populations, as well as to investigate its validity by repeating the test on three sample groups.

Method

The test consists of 14 geometric designs, each drawn on a white card 2 inches square. The first design is an equilateral triangle with sides of 1½ inches; the design on the last card is a true circle with a diameter of 1¾ inches. The 12 intermediate cards form a continuum in which the sides of the triangle become increasingly circular and finally terminate in the true circle of card 14.

The test was administered individually to three groups. The first group consisted of 45 patients in the early stages of schizophrenia, 26 females and 19 males, randomly selected from the admission and convalescent wards of the New York State Psychiatric Institute. None of these patients had been institutionalized for more than 20 months at the time of testing. The mean period of hospitalization was 4.6 months. Ages ranged from 14 to 55 years, with a mean of 25.8 years. The educational level ranged from 8 to 20 years with a mean value of 12.5 years. All patients included in this group were diagnosed as schizophrenics of various types, though the majority fell into the paranoid classification.

The second experimental group was obtained from the reception wards of Brooklyn State Hospital. They were chosen at random by the psychiatric staff from the recently admitted patients who had a diagnosis of schizophrenia. Their ages ranged from 16 to 51 years, with a mean of 31.1 years. Their educational level ranged from 6 to 16 years with a mean of 10.8 years. The 23 patients in this group comprised 13 females and 10 males.

These two patient populations may be contrasted in the following respects: the Psychiatric Institute patients were all voluntary admissions, capable of adapting to a permissive hospital milieu, of adequately caring for themselves, and in general did not pose a nursing problem. They were admitted to the hospital largely on the basis of their having a favorable prognosis. The sample drawn from Brooklyn State Hospital is more representative of the admissions usually found in a large state hospital. These patients manifest overtly bizarre behavior and many of them pose serious management problems. At least one-third of this group has a history of one or more previous hospitalizations for a mental disorder.

---

[1] This research was facilitated by a grant from the National Institute of Mental Health, Public Health Service, Department of Health, Education, and Welfare, Project M 586C. The authors are indebted to Professor Joseph Zubin, Principal Investigator, for his invaluable advice and guidance. We wish to thank Ethel Schmid and Herbert Greenwald for their kind assistance with the testing.

[2] Member of the Project staff at the time of this investigation.

A group of 23 college students, 8 females and 15 males, attending summer session courses at a large nearby university were employed as controls. These subjects were regarded as normal on the basis of their having no clinically manifest neurosis or psychosis. None were under medical or psychological treatment at the time of testing, nor had they ever been hospitalized for a mental disorder.

A test procedure consisting of the following four major parts was employed:

Part A. The 14 cards are scattered on a table before the subject, who is then requested to arrange them in one row in the best possible manner.

Part B. The true triangle and the true circle are placed at some distance from each other and the subject is requested to arrange the remaining cards (which are scattered randomly as in Part A) between these two in proper order so as to form a continuum from triangularity to circularity.

Part C. The examiner arranges the cards correctly on the continuum and the subject is requested to indicate how many cards belong in the triangle group and how many cards belong in the circle group. This establishes conceptual boundaries of a certain magnitude for these two concepts.

Part D. The subject is requested to remove all cards which are neither triangles nor circles. This measures the ability to maintain the previously established conceptual boundaries.

Results

In the absence of an alternate form of the test, and the inapplicability of the split-half method, the test-retest technique was employed to determine the dependability of the instrument over various intervening periods of time. The performances of all subjects included in the study were rated on the following parts of the test.

Part A. There are three possible levels of performance for which scores have been determined: (*a*) the arranging of the randomly placed cards along a continuum extending from triangularity to circularity or vice versa for a performance on the *highest* level; (*b*) grouping the cards in accordance with the three basic concepts of triangle, circle, and middle area, for an *intermediate* level of conceptualization; (*c*) arranging the cards as simple pairs or patterns, or with no consistency at all, for a *primitive* level. This portion of the test was not included in the computation of the reliability coefficients because of its gross susceptibility to practice effect.

Part B. A perfect score is obtained when each card is placed in its proper sequence along the continuum. A quantitative performance score is obtained by calculating the discrepancy between the expected order of the cards and the obtained order and squaring and summing these discrepancies.

Part C. This yields three scores based on the point at which the subject establishes his conceptual boundaries, C^t or the card with which the subject establishes the triangle boundary, C^c or where the boundary of the circle was placed, and C^r or the number of cards remaining in the intermediate range, being classified as neither triangles nor circles.

Part D. This gives a measure of the instability and interpenetration of boundaries established in Part C, by determining whether the boundaries C^t and C^c are maintained when the subject is requested to remove those cards which he considers to be neither triangles nor circles. This score is recorded as a "yes" or "no" entry.

The reliability coefficients for the various portions of the test and for the two groups which were retested are shown in Table 1.

Table 1

Test-Retest Correlation Coefficients for Zaslow Test Scores

Score	Normals ($N=16$)	Psychiatric Institute patients ($N=21$)
B	.14	.46
C^t	.43	.14
C^c	.90**	.33
C^r	.46*	.24
D	.42	.62†

* Significant at .05 level.
** Significant at .01 level.
† Tetrachoric *r*. The correlation of .62 for Part D was found to be significant at the .01 level if the test of significance assumes normality. When tested by a nonparametric χ^2, which does not assume a normal distribution, the relationship is significant at about the .15 level.

The mean retest interval for the normal group was 4.4 days, and for the Psychiatric Institute group it was 14.3 days.

The performance of the three sample populations is summarized in Table 2. A chi-square analysis was undertaken to determine whether any of the obtained scores were significantly differentiating. Only the first, or Part A score (performance employing high-level conceptualization) was found to be statistically significant in that a significantly higher percentage of the Normal subjects than of the Brooklyn State group employed high-level conceptualization in Part A. Also, a significantly higher percentage of the Psy-

Table 2
A Comparison of the Normal, Psychiatric Institute, and Brooklyn State Patient Groups on the Various Test Scores

Item	Normals ($N=23$)	Psychiatric Institute ($N=45$)	Brooklyn ($N=23$)
Percentage of group employing high-level conceptualization	61	67	17*
Percentage of group placing triangle between cards 3 and 5	13	38	30
Percentage of group scoring 25 or higher in Part B	0	9	13
Percentage of group placing circle at cards 11 and 12	26	42	48
Percentage of group including "middle" area of four to eight cards	26	20	30
Percentage of group maintaining boundaries in Part D	35	29	17

* Differs at the .01 level from Normals and from Psychiatric Institute group.

chiatric Institute group than of the Brooklyn State patients employed a high-level conceptualization in Part A. The difference between the Normal and Psychiatric Institute group is not statistically significant on this score. No other intergroup comparison on any of the five remaining scores yielded a statistically significant difference.

Discussion

The results of this experiment indicate that the Part B and C^t scores lacked reliability for both of the populations retested. The other three scores, Parts C^c, C^r, and D, showed statistically significant reliabilities for one or the other of the two groups, but not for both on any single portion of the test. These findings would indicate that the test as a whole is not sufficiently reliable for ordinary use.

A comparison of the three groups, with one exception, failed to yield scores on any portion of the test that would significantly differentiate the normal from the experimental groups. It was found that on Part A, the difference between the Normal and Brooklyn State sample was significant at the .01 level, but the same score did not differentiate the Psychiatric Institute group from the Normals. Since only one of six scores was found to distinguish between the experimental and control groups, we would conclude that the differential power of the instrument for the populations under comparison is very questionable.

Thus the results of this experiment indicate rather low reliability, and very questionable validity for Zaslow's test. As originally developed the test differentiated between a control and hospitalized chronic schizophrenic group. It appears to be inadequate for use in distinguishing between a normal control group and schizophrenics seen early in their hospitalization.

Summary

The Zaslow test of concept formation was administered to a group of normals, and two dissimilar groups of schizophrenics. The results were analyzed in the same manner as that used by Zaslow in his study. The findings of this study indicate that the reliability coefficients obtained from the normal and the schizophrenic group were not statistically significant for both groups on any single portion of the test. Only one of six scores was found to distinguish between the experimental and control groups. Thus, these results would indicate rather low reliability, and very questionable validity for Zaslow's test for use in differentiating between normals and hospitalized schizophrenics seen early in their hospitalization.

Received September 16, 1954.

References

1. Stacey, C. L., & Cantor, G. The use of Zaslow's test of concept formation on a group of subnormals. *J. clin. Psychol.,* 1953, **9**, 51–53.
2. Zaslow, R. W. A new approach to the problem of conceptual thinking in schizophrenia. *J. consult. Psychol.,* 1950, **14**, 335–340.

Journal of Consulting Psychology
Vol. 19, No. 3, 1955

The Effects of Serial Position and Design Difficulty on Recall of the Bender-Gestalt Test Designs[1]

Leonard D. Goodstein, Charles D. Spielberger,[2] John E. Williams,[3]
State University of Iowa
and W. Grant Dahlstrom
University of North Carolina

There is little research literature on the several "nonpersonality" variables which may affect performance on visual-motor tests. It is the purpose of this paper to offer evidence pertaining to the operation of the variables of item difficulty and order of presentation in producing differences in recalling designs from memory. It is possible that many of the differences in drawing performance among clinical groups that are currently ascribed to organic pathology or personality may be accounted for more economically within the framework of well-known serial learning principles.

In the conventional administration of the Bender-Gestalt Test (B-G), each of the nine designs is presented singly in a prescribed order and copied by the subject. A minimum of formal instruction is used and no time limits are imposed. Immediately following the last drawing, the materials are removed and the *S* may be asked to reproduce as many of the designs as he can recall (4). This technique affords a measure of the learning taking place during the copying phase of the test, and fits the paradigm of one-trial serial learning under free recall conditions, with the important exception that the *S* is not instructed to learn.

Serial position effects have been studied extensively in verbal learning situations and have been found to be an important source of variance in the recall of learned material. The usual paradigm for such studies involves successive exposures of the material until some criterion of learning is reached. Retention is then tested by some form of recall. It should be noted that the B-G procedure allows only *one* presentation of the materials with a *free* recall. No studies involving visual-motor performance approximating this paradigm have been found in the literature, and only one verbal learning study seems sufficiently similar to this procedure to serve as a guide for hypotheses concerning serial position effects on this task. Welch and Burnett (5) reported a study in which a list of nonsense syllables was presented once with instructions to learn. Testing by free recall was made immediately following the tachistoscopic presentations. The curves they reported for lists of eight syllables were interpreted as showing frequency of recall to be an increasing monotonic function of serial position in the list.

The effect of the variable of serial position upon recall can be masked by the operation of many other variables, including design difficulty. Although no experimental evidence is available as to the presence or extent of interitem differences in difficulty level on the B-G, the work of experimenters like Benton (2) strongly suggests that such differences in design difficulty on visual retention tasks may be appreciable. If this source of variance is important, it could obscure the relative contribution of serial position in determining recall frequency.

It is highly probable that many variables other than those discussed above affect the recall of the Bender designs. Age, sex, previous training, degree of maladjustment and other individual differences variables would seem to be of possible importance. The present paper is restricted to a relatively homogeneous population in terms of these variables. The major purpose of the present investigation was to study the effects of serial position and design difficulty upon the recall of the B-G designs. The experimental design employed, however, also afforded the opportunity of studying sex differences on this task.

---

[1] The authors wish to acknowledge their indebtedness to Dr. I. E. Farber of the State University of Iowa for his generous assistance in the planning and execution of this research as well as in the preparation of this manuscript.

[2] Now at Worcester State Hospital.

[3] Now at Yale University.

Experiment I

Method

Subjects. Fifty-four undergraduates, 27 males and 27 females, from an advanced undergraduate course in psychology during the Spring, 1953, semester at the State University of Iowa served as *S*s.

Procedure. In order to study the effects of serial position, nine arrangements of design presentation were used; each arrangement involved changing the serial position of the designs while holding the sequence constant.[4] Each arrangement was presented to three male and three female *S*s in an individual testing session. Each *S* was given a lead pencil and a stack of 5½ by 8½-inch white paper and then shown the designs one by one. As part of the standard instructions, *E* explained that the designs would be presented individually and that each design was to be copied on a separate page within a 40-second time limit.

Following a six-minute interpolated draw-a-person task,[5] each *S* was asked to draw as many of the original designs as he could remember, each design on a separate sheet of paper. The *E* stopped the recall session after eight minutes unless *S* stated he could recall no more designs before this time had elapsed.

Scoring. All recalled designs were scored by a single author as acceptable or unacceptable according to an arbitrary scoring scheme devised by the authors. In general, the drawings were required to be suitable in the form of the elements, free of distortions in line or dot quality, contiguity of elements, rotation, or other variations in the pattern (4). No limits were imposed on size, tone, or page orientation of the figures.[6]

Table 1

Summary of the Analysis of Variance—Experiment I

Source of variance	*df*	Mean square	*F*
Between *S*s	53		
Sex (*S*)	1	.052	
$P \times D$	8	.27	1.17
$S \times P \times D$	8	.366	1.50
Error (b)	36	.244	
Within *S*s	432		
Serial position (*P*)	8	.333	1.76
Design (*D*)	8	1.644	8.68**
$P \times D$	56	.254	1.34*
$P \times S$	8	.116	.61
$D \times S$	8	.167	.88
$S \times P \times D$	56	.179	.94
Error (w)	288	.189	
Total	485		

** $p < .001$.
* $p < .05$.

As a reliability check, the drawings were rescored by another of the authors, with 86.3 per cent agreement in scoring. Each *S*'s drawings were scored as a set, rather than scoring each design for all *S*s at one time. It is believed that this latter procedure would have improved the reliability, but the former procedure is more in line with usual clinical practice.

Results

The data, which consisted of the proportions of correctly recalled designs, were analyzed by a Lindquist Type IV mixed design (3, pp. 285–288) in which the separate B-G designs (*D*) and serial positions (*P*) comprised the Latin square factors and sex (*S*) was a between-*S*s factor. In this experimental design, the individual designs and the serial positions are completely counterbalanced. Essentially, the experiment involved two replications of the counterbalanced factors, one with male *S*s and the other with female *S*s.

The results of the analysis of variance of these data are reported in Table 1. As can be

---

[4] A–1–2 . . . 8, 8–A–1– . . . 7, and 7–8–A– . . . 6 have the same sequence of items, but each design occurs in a different serial position.

[5] This study is concerned only with the recall drawings. No analysis of the original drawings was made. The draw-a-person task was used only to provide a uniform means of filling the interval between the original drawing and recall.

[6] Mimeographed copies of these scoring standards may be obtained upon request.

Table 2

Proportion of *S*s Correctly Recalling Each Design and the *SE* of Each Proportion

Measure	Designs A	1	2	3	4	5	6	7	8
Proportion	.85	.89	.80	.41	.50	.67	.63	.44	.65
SE prop	.05	.04	.05	.07	.07	.06	.07	.07	.07

seen, no significant effect was found attributable to sex. Therefore, for subsequent analyses, the data for the two sexes were combined.

The obtained *F* for design difficulty (*D*) was highly significant ($p < .001$). Table 2 reports the proportion of *S*s correctly reproducing each design and the *SE* of each proportion. The differences among the proportions for the individual designs suggest that there are three levels of difficulty with designs A, 1, and 2 about equally easy; designs 5, 6, and 8 about equal at an intermediate level of difficulty; and designs 3, 4, and 7 the most difficult according to the scoring scheme used.

The *F* test for serial position (*P*) in Table 1 attained only a moderate level of statistical significance ($p < .08$). The serial position effects presented in Fig. 1, however, are so similar to the results obtained by Welch and Burnett (5) that further research seemed warranted.

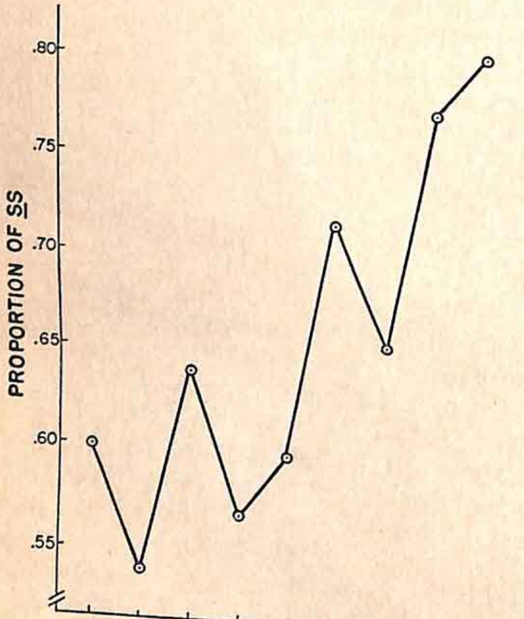

Fig. 1. The proportion of *S*s correctly recalling designs at each serial position in Experiment I. The individual designs were presented equally often in each position counterbalancing the effect of design difficulty. The data from men and women were pooled.

Although the obtained *F* for the within component of the "mixed" $P \times D$ interaction was barely significant ($p < .05$), the hypothesis of no interaction cannot be rejected in this type of experimental design unless either or both components of this interaction are significant at the .025 level of confidence (3, p. 280). In the present experiment, therefore, the hypothesis of no interaction must be retained.

The obtained results suggested that the recall of a design might be affected not only by the difficulty level of the design but also by its serial position in a given order of presentation; that is, design difficulty might interact with order of presentation. To test this hypothesis, as well as to cross validate the results as to sex and design difficulty, a second experiment was performed.

Experiment II

Method

Procedure. Two orders of presentation were developed on the basis of the rough estimates obtained from Experiment I of the relative difficulty of the nine designs (Table 2) and of the nine serial positions (Fig. 1). Order I (3, 7, 4, 6, 8, 5, 2, A, 1) involved placing the most difficult designs in the most difficult (the initial) serial positions and the easiest designs in the easiest (the final) positions. Order II (1, A, 2, 5, 8, 6, 4, 7, 3) placed the most difficult designs in the easiest positions and the easiest designs in the most difficult positions. These orders permitted testing the interaction between difficulty of the designs and order of presentation. Each order was presented to 15 male and 15 female *S*s. All other procedures were identical to those employed in the first experiment.

Subjects. Sixty undergraduates, 30 males and 30 females, from the introductory psychology course during the Fall, 1953, semester at the State University of Iowa served as *S*s.

Scoring. All recalled designs were scored independently by three of the authors, according to the same scoring scheme used in Experiment I. The scoring of an individual design was the score given by the majority of the judges. The percentage of agreement for all three judges was 84.3; for the three possible pairs of scorers the percentages of agreement were 91.8, 89.7, and 87.1.

Results

The proportions of hard and easy designs correctly recalled under the two orders of presentation are shown in Fig. 2. The recall of designs appears to be a function of both design difficulty and order of presentation, that is, serial position. There was little difference between the proportions of easy designs (A, 1, 2) correctly recalled at the hard (1–3) and easy (7–9) serial positions. The hard designs (3, 4, 7), however, were correctly recalled by only 38 per cent of *S*s when they appear in the hard positions, while in the easy positions they were recalled by 58 per cent of *S*s. Order of presentation appears to alter the recall of the hard items but has little or no effect on the easy items.

These data were analyzed by a Lindquist Type III mixed design (3, pp. 281–284) in which sex (*S*) and the order of presentation (*O*) were between-*S*s factors and the separate B-G designs (*D*) with a within-*S*s factor. The results of this analysis are reported in Table 3.

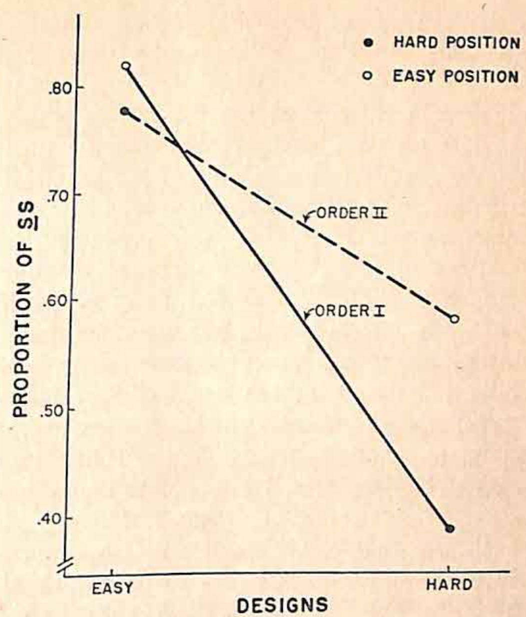

Fig. 2. The proportion of *S*s correctly recalling the hard and easy designs in the two orders of presentation. These curves are based upon the averages for the easiest serial positions (7–9) and the most difficult serial positions (1–3). While only data from the hard (3, 4, 7) and easy (A, 1, 2) designs are presented here, the data from all designs were included in the statistical analysis. The data from men and women were pooled.

As can be seen, the obtained *F* for the interaction between design and order of presentation was highly significant ($p < .001$), indicating that the recall of a given design is a function of *both* its difficulty level and its serial position in a given order of presentation. No other interactions were statistically significant.

Table 3

Summary of the Analysis of Variance—Experiment II

Source of variance	*df*	Mean square	*F*
Between *S*s			
Sex (*S*)	1	.011	
Order (*O*)	1	.010	
$S \times O$	1	.153	
Error (b)	56	.201	
Within *S*s	480		
Design (*D*)	8	2.860	16.22**
$D \times O$	8	.451	2.56**
$D \times S$	8	.245	1.39
$D \times O \times S$	8		
Error (w)	448	.176	
Total	540		

** $p < .001$.

The obtained *F* for design difficulty was the only significant ($p < .001$) main effect, suggesting that difficulty level is the more important variable affecting recall. No significant effect was found attributable to sex differences; therefore, in all the other analyses the data for the two sexes were combined. These results clearly confirm those of Experiment I.

Discussion

The procedure of this study deviated from the conventional clinical use of the B-G largely in the direction of greater precision and uniformity of stimulus presentation. Thus, the time spent on each design was held constant, the design exposures were further controlled by the use of individual sheets for copying each of the designs, and the instructions were standardized. The interpolation of an interval filled with a drawing task was calculated to make the recall task more difficult. This was done in order to get greater variability in the frequency of designs recalled in this group with high intellectual ability and to control the manner in which the interval was filled. Only the accuracy of the recalled designs was examined in the present study while clinical usage usually involves many more features of the drawings (4).

It seems clear that these supposedly simple designs do differ in the frequency of recall even in this academically superior group. Little allowance is made for these design difficulty differences in the usual clinical interpretations of these drawings. Conclusions about a particular set of recall drawings should be tempered by adequate consideration of the variable of design difficulty.

In the clinical use of the B-G a conventional order (A–1–2– . . . 8) is employed. In this order the easiest designs are in the most difficult serial positions while the most difficult designs are in positions of intermediate difficulty. It is not inconceivable that modifications in the conventional order might produce a more effective psychodiagnostic instrument.

The absence of any significant interaction effects between design difficulty and sex casts doubt on many of the clinical hunches which relate sexually relevant personality characteristics to particular design reproductions. The scoring for acceptability, however, may not be sufficiently sensitive to such distortions.

Summary and Conclusions

In an attempt to provide a firmer experimental foundation for the use of visual-motor tasks in the clinic, the effects of order of presentation, design difficulty, and sex on the free recall of the designs on the Bender Visual-Motor Gestalt Test (B-G) were investigated.

In Experiment I 54 college *Ss*, equally divided as to sex, were presented with one of the nine orders of the B-G designs, followed by a draw-a-person task and a free recall session. The recalled B-G designs were scored for accuracy and the frequency of correct recall was treated by an analysis of variance for design difficulty, serial position, and sex. The effect of design difficulty was highly significant, while the effect of serial position reached a low level of statistical significance. The design difficulty × serial position interaction effects were equivocal and difficult to interpret.

In Experiment II 60 college *Ss*, equally divided as to sex, were presented with two orders of the B-G designs. In one order the most difficult designs were placed in the most difficult serial positions while in the other order the most difficult designs were placed in the easiest serial positions; difficulty of designs and of serial position was estimated from Experiment I. The remainder of the procedure was identical to that of Experiment I. The designs were scored for accuracy and the frequency of correct recall was treated by an analysis of variance for design difficulty, order of presentation, and sex. The effects of design difficulty and the interaction between order of presentation and difficulty were the only statistically significant effects.

The present investigation would seem to support the following conclusions:

1. The B-G designs differ significantly in ease of recall, even in an intellectually superior population.
2. B-G designs A, 1, and 2 appear to be the easiest; designs 3, 4, and 7 appear to be the most difficult; the remaining designs fall at an intermediate level of difficulty.
3. The recall of a given design is a function of both its difficulty level and its serial position in a given order of presentation.
4. At the college level, males and females do not differ in total frequency of correct recall, nor in frequency of correct reproduction of any particular design.

Received October 22, 1954.

References

1. Bender, Lauretta. A visual motor Gestalt test and its clinical use. *Res. Monogr. Amer. Orthopsychiat. Ass.*, 1938, No. 3.
2. Benton, A. L. A visual retention test for clinical use. *Arch. Neurol. Psychiat.*, 1945, **54**, 212–216.
3. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
4. Peak, R. M., & Quast, W. *A scoring system for the Bender-Gestalt.* Hastings, Minn. (Box 292) and Minneapolis, Minn. (2810 42nd St.): Authors, 1951.
5. Welch, G. B., & Burnett, C. T. Is primacy a factor in association formation? *Amer. J. Psychol.*, 1924, **35**, 396–401.

NEW BOOKS AND TESTS

Books

Allen, Robert M. *Elements of Rorschach interpretation.* New York: International Universities Press, 1954. Pp. v + 242. $4.00.

Although this textbook on the use of the Rorschach is presented at a rather elementary level, it can scarcely be claimed that the *elements* of Rorschach interpretation receive a comprehensive treatment. Taking as his theme the statement that "every performance of a person is an expression of his whole personality, perception included," the author proceeds to offer personality interpretations for various modes of response to the Rorschach cards. In general, his discussions are rich in claims for the significance of Rorschach responses, but poor in rational and empirical justification for these claims. To a conspicuous degree, the textbook is a recital of the point of view, opinions, insights, and practices of the author. It does not appear to be an impersonal presentation of commonly accepted elemental bases for Rorschach interpretations.—*J. R. W.*

Deutsch, Felix, & Murphy, William F. *The clinical interview.* Vol. 1. *Diagnosis.* New York: International Universities Press, 1955. Pp. 613. $10.00.

This is the first of two volumes on the clinical interview, and is devoted to diagnosis; the second volume, promised shortly, is to deal with the clinical interview and psychotherapy. Both are designed for the training of psychiatric residents. Judging from the published volume, they can also serve well in the training of clinical psychologists. The authors develop two interesting concepts: the "associative anamnesis," which is a method of "guided free association" for diagnostic purposes, and "sector psychotherapy," which is a goal-limited type of therapy. By means of 21 recorded interviews with different patients showing different syndromes, Deutsch and Murphy convey their technique in detail. Each interview is preceded by a discussion of the syndrome involved and a brief summary of the case history; the interviews are generously interlarded with comments and interpretations. Each interview is designed as a complete entity, and is expected to achieve a predetermined goal. A chapter on "The Application of Psychoanalytic Concepts to Resident Training and the Technique of the Therapeutic Interview" is particularly worthwhile.—*M. K.*

Donahue, Wilma T. (Ed.) *Education for later maturity: a handbook.* New York: Whiteside, Inc. & William Morrow, 1955. Pp. xiii + 338. $4.50.

Our machine age has produced increasing hours of leisure. At the same time there has been a rapid expansion of the life span, so that the proportion of older people in our society has increased. Adult education finds a real challenge in meeting the needs and demands of these mature years. *Education for Later Maturity* presents the problems of the aging period, together with the present attempts to meet them. It stresses the guiding principles that must be established in order to make satisfactory solutions. Adult education is envisaged not as a meeting of emergency needs, but as a lifelong process. If so conceived, the well adjusted years of maturity are not a period requiring specific education, but are the culmination of a process of lifelong adjustment to growth and change. Each chapter presents a separate aspect of the problem by different authors. The book is so carefully conceived and edited that there is no overlapping. The reader has a feeling of unity and cohesiveness unusual in a book of this nature. There is excellent coverage of all phases of adult education, and a careful review of the present program, often with a frank evaluation of possibilities and limitations. The book will prove profitable reading for anyone who is working with mature people or who is facing the problems of later maturity himself.—*B. M. L.*

Edwards, Allen L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954. Pp. vii + 542. $6.50.

This is an elementary book in statistics. With respect to style, format, and content it falls clearly in the pattern of textbooks that have been accepted by departments of psychology for undergraduate courses in statistics. The area covered, however, exceeds the work that can be expected from a one-semester course at the undergraduate level. For the most part the treatment is modern and it should be noted that a chapter on significance tests for ranked data is included. It is likely that the reader will find this text to be an improvement over Edwards' earlier text on statistical analysis—*J. R. W.*

Note.—The reviews were prepared by the Editor and the Advisory Editors, who may be identified by their initials.

Macfarlane, Jean W., Allen, Lucille, & Honzik, Marjorie P. *A developmental study of the behavior problems of normal children between twenty-one months and fourteen years.* Berkeley and Los Angeles, Calif.: Univer. of California Press, 1954. Pp. vii + 222 (paper). $2.25.

This report is offered as a contribution to our knowledge of the problems and developmental circumstances of children who are not the subject of special therapy or guidance. The writers suggest that the data are more descriptive of a "normal" population of families than are data based on families who come to clinics because of referrals. Although the work presumes to offer no more than a description and suffers the obvious limitations of an account of children who lived in one particular area (and of necessity at one particular time), it is a storehouse of information and may be a source of hypotheses for subsequent investigations and a catalog of bench marks useful for comparison and reference. The sample comprised 126 infants who were examined at regular subsequent intervals, with diminution of numbers resulting from the usual factors which interfere with "follow-up" studies. One of the most interesting features of this publication is the presentation of the relative incidence of various types of "problem" behavior through the critical period beginning at age five and ending at age fourteen.—*J. R. W.*

Meerloo, J. A. M. *The two faces of man.* New York: International Universities Press, 1954. Pp. x + 237. $4.00.

This little volume contains two essentially independent long essays. One, entitled "Father Time," is a thoughtful consideration of the psychology of time perception and the experience of duration and temporal relationships. The other, the title essay, is concerned with ambivalence and the psychological significance of the bipolarities and antinomies that seem omnipresent in the world of experience. In both essays, Meerloo demonstrates a rich background of clinical experience and a sensitive familiarity with the tensions that arise from the ambiguities of complex social life. His erudition is considerable, and his leisurely, learned pace sometimes masks a boldness of thought disciplined by a genuine feeling for evidence and logic in the testing of intuitive formulations. Psychologists will find much in this slim book to sharpen their clinical insights and to challenge their investigative resourcefulness. A Dutch psychiatrist now practicing in the United States, Dr. Meerloo is known through his previous books for his ingenuity in psychological conceptualization and his easy, civilized English style. *The Two Faces of Man* continues and justifies his reputation for innovation and gentleness.—*E. J. S.*

Remmers, H. H. *Introduction to opinion and attitude measurement.* New York: Harper, 1954. Pp. viii + 437. $5.00.

Although written as a text for an undergraduate course in attitude measurement, this well-written volume provides a useful survey of the methods and problems of attitude measurement for anyone whose training did not include such a course at either the undergraduate or graduate level. Part I, entitled "Techniques of Attitude and Opinion Measurement," includes a discussion of all current techniques except those recently introduced by Coombs. Part II, the last half of the book, includes chapters on the application of opinion and attitude measurement in business, government, industry, community relations and education. Each chapter is concluded with a series of thought-provoking questions (many of which are not answered in the text!) and a selected bibliography.—*E. L. K.*

Robinson, Mary F., & Freeman, Walter. *Psychosurgery and the self.* New York: Grune & Stratton, 1954. Pp. ix + 118. $3.00.

There is packed into this clear and carefully written 118-page book an explanation of psychosurgery including the less drastic transorbital lobotomy, descriptions of postlobotomy personalities, an objective and concise review of the literature on the procedure as well as current psychological concepts of the self. This gives perspective to an experiment on the effects of the standard psychosurgery on the patient's self as compared with a control group of patients who recovered without the operation. Dr. Robinson's basic hypothesis is that psychosurgery changes the structure of the self through reducing the capacity for the feeling of self-continuity. Her design provides evidence substantiating the hypothesis, and her experimental adventure into this important but unexplored area of psychology should stimulate theory and research. Her methods seem intriguingly simple but penetrating, and her insights into the fundamental changes after prefrontal operations are satisfying. The undesirable changes which are found along with reduced anxiety, introversion, and neuroticism are indicated throughout the book, and it is clearly stated that the study is investigative rather than evaluative. Nevertheless, the reviewer feels that, despite Dr. Robinson's scientific objectivity and honesty, the flavor of the volume is definitely favorable to psychosurgery.—*F. McK.*

Sappenfield, Bert R. *Personality dynamics.* New York: Alfred A. Knopf, 1954. Pp. xiv + 412 + xvi. $5.50.

In this text, the subtitle of which is "An Integrative Psychology of Adjustment," the author has attempted, first, a systematic presentation of psychoanalytic principles and, second, an integration of psychodynamic principles with an organismic conception of behavior. Furthermore, although employing much of Freud's original terminology, the author attempts to redefine such terms in language more common to American psychology (the glossary contains almost 300 such definitions). In these efforts at integration, the author proposes what he regards as four contributions to conceptual clarification: (*a*) the definition of id functions as those involving biogenic need-tensions, ego functions as those involving cognitive (intellectual) and voluntary processes, and superego functions as involving all aspects of psychogenic motivation; (*b*) the analysis of motives into three components: need, instrumental act, and object; (*c*) equating the concepts "id functions" and anxiety and defining both as "the occurrence in consciousness of biogenic need-tensions without conscious representation of instrumental acts or objects in relation to which gratification or relief may be anticipated"; (*d*) the distinction between perceptual identification and developmental identification. Emphasis throughout the text is on the development and functioning of the normal personality, and although pathological conditions are often used by way of illustration, it is not a textbook in abnormal psychology. For example, the word "schizophrenia" does not appear in a twelve-page subject index. Instead, it is a text which seeks within a single volume (*a*) to familiarize the student with all essential aspects of psychoanalytic theory, and (*b*) to get him in the habit of using this theoretical formulation and terminology in thinking about his own behavior and that of his peers. As such it is not an easy text. I predict that undergraduates will find it interesting but difficult. Graduate students should find it a helpful introduction, especially in view of the many references to original sources. The style is almost completely didactic with practically no reference to the methodological problems of assessing dynamic processes, to the variables involved, or to empirical studies bearing on the topics discussed.—*E. L. K.*

Savage, Leonard J. *The foundations of statistics.* New York: Wiley, 1954. Pp. xv + 294. $6.00.

Lest the prospective reader be misled by the title, it should be stated at once that this is not an introductory text in statistics. In fact, it is not a textbook at all but rather a highly sophisticated discussion of the philosophical and logical bases of statistics "as a discipline of rational decision in the fact of uncertainty." The author develops and defends the currently somewhat unfashionable "personalistic viewpoint" of probability and then discusses its implications for observation and experimentation, for decision making, and for current statistical theory and practice. The preface suggests that any but the casual reader should have some formal preparation in the theory of mathematical probability, at least a year of calculus and some training in formal logic. In addition, it is suggested that he should "sit bolt upright on a hard chair at a desk." Even though the present reviewer does not pretend to have worked his way through all of the book in this manner, he did find the less mathematical chapters highly informative and thought provoking. He believes that many other psychologists will find it worth a few hours of their time. A bibliography of some 170 items is included.—*E. L. K.*

Tests

Drake, Raleigh M. *Drake Musical Aptitude Tests.* Ages 8–adult. 2 forms. (80) min. Microgroove phonograph record ($5.95); self-scoring answer pads, musical memory ($2.45 per 20), rhythm ($2.45 per 20); manual, pp. 32 (75¢); specimen set ($6.95). Chicago: Science Research Associates, 1954.

The Drake tests are the outcome of a life career devoted to measurement in music, and are a substantial contribution to their field. In place of the more familiar potpourri of short samples, these tests measure only two kinds of performance, at a length that permits adequate reliability (mainly .85 to .95). The two subtests have very low correlations with each other and with age and intelligence, and moderate correlations (about .35) with musical training. The norms are remarkably stable among several national and ethnic groups. Validity, unfortunately, is reported only in terms of correlations with teachers' ratings, but such correlations are about as high as the reliabilities of the ratings would permit. The manual conforms to high standards in displaying extensive and relevant data. Both musically and psychometrically, this is a good test.—*L. F. S.*

Thorpe, Louis P., Meyers, Charles E., & Sea, Marcella R. *What I like to do, an inventory of children's interests.* Grades 4–6. 1 form. Untimed, (50–60) min. Hand or IBM scoring. Booklet ($2.95 per 20); IBM answer sheet ($2.90 per 100); profile sheet (90¢ per 20); manual, pp. 16 (25¢); specimen set (50¢). Chicago: Science Research Associates, 1954.

What I Like To Do is an uninspired but workmanlike job in the construction of an interest inventory for grades 4 to 6 and, by implication but without data, for grade 7. The 294 items assess interests in art, music, social studies, active play, quiet play, manual arts, home arts, and science. The items were developed from the literature on children's interests, and were screened both by judgments and by an empirical trial. National norms are based on 3,803 children in grades 4 to 6, and Kuder-Richardson reliabilities of the separate parts center around .87. A general factor of liking things seems to be present, as most of the scale intercorrelations exceed .50.—*L. F. S .*

PSYCHIATRY

Journal for the Study of Interpersonal Processes

CONTENTS OF NO. 1, VOL. XVIII, FEBRUARY 1955

GEORGE H. DESSION: *Psychiatry and Public Policy*

MARTIN GROTJAHN: *Problems and Techniques of Supervision*

LIONEL OVESEY: *The Pseudohomosexual Anxiety*

R. RICHARD WOHL and HARRY TROSMAN: *A Retrospect of Freud's* Leonardo

DONALD MELTZER: *Toward a Structural Concept of Anxiety*

SIMON O. LESSER: *The Functions of Form in Narrative Art*

HELEN CODERE: *A Genealogical Study of Kinship in the United States*

HANS H. TOCH and ALBERT H. HASTORF: *Homeostasis in Psychology*

EDITORIAL NOTES.

BRIEF COMMUNICATIONS.

BOOK REVIEWS.

Published in February, May, August, and November by the

William Alanson White Psychiatric Foundation

1711 Rhode Island Ave., N. W., Washington 6, D. C.

Subscription price $10.00 for 1955; foreign $10.80
Single copy $2.50; foreign $2.70

JOURNAL OF
CONSULTING
PSYCHOLOGY

August, 1955 Vol. 19, No. 4

Contents

Changes in Psychoneurotic Patients With and Without Psychotherapy: *Frank Barron and Timothy F. Leary* - 239

Brief Psychotherapy and Enuresis: *Daniel H. Harris, Richard W. Firestone, and Carl M. Wagner* - 246

The Measurement of Depth of Interpretation: *Norman I. Harway, Allen T. Dittmann, Harold L. Raush, Edward S. Bordin, and David Rigler* - - - - - - - - - - - 247

Responses to Imagined Auditory Stimuli as Compared to Recorded Sounds: *David R. Stone* - 254

Clinical Statements as Scientific Propositions and Social Decisions: *Frank M. du Mas* - - 255

Delayed Gratification and Ego Development: Implications for Clinical and Experimental Research: *Jerome L. Singer* - 259

Testing Behavior Hypotheses with the Rorschach: An Exploration in Validation: *Arthur E. Eschenbach and Edgar F. Borgatta* - - - - - - - - - - - - - - - - - - 267

Predicting Success on the Member-Employee Rehabilitation Program: *Bernard A. Stotsky* 274

Relations Among Several Objective Measures of Anxiety Under Different Conditions of Motivation: *Anthony Davids* - 275

A Further Investigation of the Relationship Between Manifest Anxiety and Intelligence: *Allen D. Calvin, Paul B. Koons, Jr., Joseph L. Bingham, and Howard H. Fink* - - 280

The Correlation of Intelligence and Neurological Findings on Twenty-Two Patients Autopsied for Brain Damage: *Robert S. Morrow and Joseph C. Mark* - - - - - - 283

The Awareness of Hostile Feelings in Patients by Their Nurses: *Sol Charen* - - - - - - 290

Psychological Changes over a Three-Year Period Following Bilateral Prefrontal Lobotomy: *Isidor W. Scherer, John F. Winne, and Robert W. Baker* - - - - - - - - - - - 291

Diagnosis of Organicity by Means of Spiral Aftereffect: *A. Cooper Price and H. L. Deabler* 299

The Efficacy of Diagnostic Pattern Analysis with the Wechsler-Bellevue: *Jacob Cohen* - - 303

Specific Ability in IQ Change: *Charles T. Baker, Lester W. Sontag, and Virginia L. Nelson* 307

Rigidity on the Rorschach and Response to Intermittent Photic Stimulation: *Laverne C. Johnson and John A. Stern* - - - - - - - - - - - - - - - - - - - 311

New Books and Tests - 318

Bureau Edn. Psy. Research
[illegible] HA - - [illegible] COLLEGE
Dated
Accs No

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 70 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 70 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 70-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. —— from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $—— for microfilm or $—— for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Extended report. The full report is prepared in the style specified by the *Publication Manual* (1), except that it may be typed with single spacing for economy in photoduplication by the ADI.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.*, 1952, **49**, 389–449.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Changes in Psychoneurotic Patients With and Without Psychotherapy[1]

Frank Barron
Institute of Personality Assessment and Research, University of California, Berkeley

and Timothy F. Leary
Kaiser Foundation Hospital, Oakland, California

Two tendencies seem constantly at work in living systems: the tendency to integrate a durable and stable arrangement of internal parts and external relationships, and the interacting opposite of this, the tendency to respond variably either to pressures arising from internal structure or to new external conditions in such a way as to break down old forms of organization and replace them with new ones.

In psychoneurosis, which is in part at least an attempt to stabilize matters internally and externally, one might expect also that forces arising internally and tending towards change would be at work. As Eysenck (3) has pointed out in his review of reports on the incidence of improvement under varying therapeutic regimes, the rate of spontaneous remission in untreated psychoneuroses is almost identical with the recovery rate reported when intensive psychotherapy has been applied. While Eysenck's analysis of these reported data has been challenged by Rosenzweig (9), and perhaps cannot be considered to be compelling in the light of Rosenzweig's criticisms, it is undoubtedly true that a substantial percentage of untreated psychoneuroses and psychoses make some sort of recovery.

To say that a remission is spontaneous, however, is generally to make a confession of ignorance; what we mean is that a change has occurred for reasons which we do not understand. When a change in the psychic state of a patient has occurred concurrently with the application of psychotherapy, we usually feel that we *do* understand the causes of the change. Unless, however, we are certain that there is a significant increment in the recovery rate when psychotherapy is applied, we may be deceived in attributing the changes we have observed to the psychotherapeutic forces we think we have applied. Hence the importance of research into psychological change under systematically varied conditions.

The Present Research

This paper reports the results of a study in which 150 psychoneurotic patients all drawn from the same clinic population were tested with the Minnesota Multiphasic Personality Inventory before and after an interval of time during which some of them received psychotherapy and some of them did not. All 150 patients had applied for psychotherapy in the psychiatric clinic in which the study was conducted, and all of them had been accepted for treatment; a group of 23 cases, however, had to be placed on a waiting list until therapeutic facilities became available, and they waited some six months before beginning treatment. Of those who received psychotherapy immediately, 85 received group therapy, and 42 received individual therapy; they included only those patients treated for a minimum of three months during the period of time covered by the study.

---

[1] The present investigation is one of a series of studies sponsored by the Kaiser Foundation, Oakland, California, and supported in part by Research Grant MH-331 from the National Institute of Mental Health, Public Health Service. Part of the financial support for the present study came from a grant of the Medical Division of the Rockefeller Foundation to the Institute of Personality Assessment and Research, Berkeley, California.

The aim of the study is to discover what changes, if any, occur in MMPI profiles of patients who receive psychotherapy, and to compare these changes with whatever changes may be observed in the MMPI profiles of patients who during the same period of time remain untreated.

Other investigators, such as Kaufmann (5), Gallagher (4), Mosak (7), and Rashkis and Shaskin (8), have reported observations of MMPI changes following psychotherapy; the distinctive feature of the present research is the use of a control group consisting of patients who are in need of therapy and who want it, but who do not receive it.

The psychotherapists who worked with the treatment groups included psychiatrists, social workers, and psychologists. All therapists had at least three years of postdoctoral or postgraduate training and experience. The theoretical orientation in individual therapy was psychoanalytic, although the psychoanalytic method was not used. The explicit therapeutic strategy or time perspective for individual treatment varied from patient to patient, but what occurred generally was brief, ego-oriented psychotherapy on a once-a-week basis. Group therapy differed somewhat in its greater emphasis upon current interpersonal responses; i.e., the interpretations and focusing remarks of the therapist were centered on the interpersonal tactics of the patient in relation to the therapist and the five other patients in the group, the reactions he provoked from other group members, and the meaning of these interactions in terms of the patient's total life situation.

The Research Situation

This study was done in the psychiatric outpatient clinic of the Kaiser Foundation Hospital in Oakland, California, and was conducted with a minimum of interference with the usual clinic routine. No special arrangements were made for the assignment of patients to treatment or nontreatment groups, and neither patients nor psychotherapists considered themselves involved in a research program.

This is perhaps not the ideal design for a study in which an untreated control group is to be used to establish a base line for evaluating the unique contribution of psychotherapy to the amelioration of the psychoneurosis. In a less real world, for example, the research psychologist in a clinic might assign new patients randomly to treatment and nontreatment groups, and he might also set some standard interval as the period of time which should elapse between the determination of the patient's state at the beginning and at the end of the experiment. The experimental (treatment) and control (nontreatment) groups would then be comparable in terms of principle of selection, and might be expected not to differ significantly from one another in severity of initial condition, diagnosis, or prognosis. Furthermore, an equal length of time would have elapsed between $t-1$ (initial state) and $t-2$ (terminal state) for both groups.

Under the pressing practical demands of clinical work, it is difficult to arrange a research design which will satisfy perfectly these ideal conditions. The decision not to treat is always a critical one in any therapeutic practice, and in psychiatric work especially such risks as suicide, psychosis, and various other personal and interpersonal catastrophes are possibly attendant upon an unfortunate decision. Random selection of patients for treatment and nontreatment groups is therefore not possible in a situation where the medical responsibility is properly the paramount one.

It is also not possible to establish in advance the exact duration of treatment. Treatment may be terminated at the decision of the patient at any time. Furthermore, early termination of treatment is sometimes indicated for technical reasons, and is brought about by the therapist. Again, in some cases the tempo of work is much slower, and treatment necessarily takes longer. Finally, accidents of circumstance, such as a change in residence or in place of employment, may operate to cause interruption of treatment before the lapse of time specified as desirable for the experiment.

In spite of these irregularities, the design of the present study remains fairly clear-cut.

As it happened, the experimental and control groups turned out to be quite comparable with one another in diagnosis, prognosis, severity of initial condition, age, sex, and education. These comparisons will be presented below. Furthermore, the interval between $t-1$ and $t-2$ was an average of 7.00 months for the nontreatment group, 8.32 months for the individual psychotherapy group, and 8.63 months for the group therapy group. These means do not differ significantly from one another (F of 2.5, $p > .05$).

Characteristics of the Sample

None of the patients in this study was so ill at the time of application for treatment as to require hospitalization. They were diagnosed most commonly as psychosomatic neuroses, obsessive or schizoid characters, or phobics (anxiety neuroses, etc.); less commonly, as psychopathic characters or as hysterics. The average age of the group was 32 years, the average education was one year of college. Of the 150 patients, 97 were women; this preponderance of women is characteristic of the clinic.

The difference in proportions of men to women in treatment and nontreatment groups was not statistically significant: women constituted 67 per cent of the nontreatment group and 63 per cent of the treatment group, a percentage difference which gives a *CR* of 0.37, $p > .05$. No sex differences were found in any of the further comparisons presented in this paper, though computations were done separately for men and women; combined totals for both sexes are therefore given when statistics are cited.

It was indicated above that treatment and nontreatment groups did not differ significantly in age, education, or diagnosis. More exactly, they compared as follows: In age, the mean of the treatment group was 32.48, that of the nontreatment group 30.9; the sigmas were 6.8 and 5.7, respectively. In years of schooling, the mean of the treatment group was 13.6, that of the nontreatment group was 12.8; the respective sigmas were 2.7 and 2.3. In diagnosis, the group percentages in the types of neuroses recognized in the simplified, MMPI-based diagnostic scheme (6) used in the Kaiser Foundation Psychological Research Project were as follows (treatment group percentage given first, nontreatment group next): Psychosomatic 19 per cent, 18 per cent; Hysteric 6 per cent, 9 per cent; Psychopathic 7 per cent, 5 per cent; Obsessive 19 per cent, 23 per cent; Schizoid 21 per cent, 23 per cent; Phobic 12 per cent, 18 per cent; Unclassified 16 per cent, 5 per cent. None of the percentage differences in any category is statistically significant; the largest *CR*, that for the difference in the unclassified category, was 1.54, the others being less than 1.00.

The patients in the treatment group who were assigned to individual psychotherapy were also compared as a class with those patients assigned to group psychotherapy. These two treatment subgroups did not differ significantly from one another in any of the comparisons presented above. Hereafter the descriptive statistics for the two subgroups will be presented in detail, since it is of some interest to compare the amount and kind of change following group psychotherapy with the amount and kind of change following individual psychotherapy.

The logic of the experiment requires that treatment and nontreatment groups be comparable in terms of prognosis and of severity of initial condition, as well as of diagnosis. The MMPI provides a basis of comparison on both points. The test was scored not only on the commonly used clinical and validity scales, where scores would indicate severity of the psychoneurosis, but also on a new scale developed by one of the authors (1). This scale, called the Ego-strength (*Es*) scale, was designed originally to predict response of psychoneurotic patients to psychotherapy, and it proved to have substantial validity for that purpose in three samples beyond the standardization sample. Its other correlates in both normal (2) and clinic samples led to its being interpreted as a measure of ego strength, and it was accordingly so named; however, its most important clinical function is as a prognostic indicator.

On all of these MMPI scales, then, there should be no difference between the treatment and the nontreatment groups at $t-1$. The appropriate comparisons are presented in Table 1.

Table 1

Comparison of Treatment and Nontreatment Groups on MMPI Scales at $t-1$

MMPI scale	Individual psychotherapy ($N=42$)		Group psychotherapy ($N=85$)		No treatment ($N=23$)		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	*F*†
1. Lie	48.04	6.76	47.70	5.25	49.70	5.19	1.03
2. *F*	61.00	12.01	60.14	12.16	58.74	7.94	0.28
3. *K*	51.16	7.29	51.34	8.31	54.96	4.89	2.25
4. *Hs*	61.40	13.61	61.18	13.19	60.04	15.43	0.08
5. *D*	78.71	12.25	73.60	13.72	76.13	12.92	2.08
6. *Hy*	67.21	8.87	66.40	10.38	66.87	13.78	0.08
7. *Pd*	67.57	14.65	66.75	12.08	67.83	12.73	0.09
8. *Pa*	63.66	10.22	61.17	10.76	62.13	8.37	0.82
9. *Pt*	73.09	13.75	69.52	12.70	72.65	12.00	1.29
10. *Sc*	71.50	17.56	68.00	13.69	67.83	12.81	0.88
11. *Ma*	55.88	10.51	54.80	10.32	49.96	10.05	2.84
12. *Es**	41.05	3.46	41.76	7.35	40.80	6.16	0.24

* Raw scores rather than *T* scores are given for *Es*. The mean for clinic populations is 41, and the *SD* 7.5.

† With *df*'s of $n_1 = 2$ and $n_2 = 147$, the *F* ratio must be 3.00 or greater to be significant at the .05 level of confidence. None of the *F* tests shown here indicate statistically significant differences.

As Table 1 shows, the groups did not differ significantly from one another either in initial severity of condition as indicated by the MMPI clinical scales, or in prognosis, as indicated by the *Es* scale. Since they were similar also in diagnosis, and in age and education, it may be concluded, as far as the main variables relevant to this study are concerned, that all three subsamples were drawn from the same population.

Since this is so, the essential conditions on which the groups are varied are these: (*a*) whether or not the patients received psychotherapy in the interval between $t-1$ and $t-2$; and (*b*) whether the patients who received psychotherapy were given individual or group treatment. The relationships between these conditions and the quantity and kind of change, as shown in the differences between MMPI profiles at $t-1$ and $t-2$, will be described next.

Change with Psychotherapy

For each MMPI scale, the discrepancy between the scale score at $t-1$ and at $t-2$ was computed for each patient. If no change at all had been brought about by psychotherapy, one would expect the mean discrepancy scores on all MMPI scales for the treatment groups to vary in a chance manner about zero. That such is not the case may be seen from Table 2.

As Table 2 shows, a significant decrease occurs in the Depression, Hysteria, Hypochondriasis, and Lie scores following psychotherapy, whether individual or group. With group psychotherapy, there is a significant decrease in the Paranoia and Psychasthenia scores as well. With both forms of therapy, there is a significant rise in the Ego-strength scores. Finally, for the individual psychotherapy patients there is also a significant rise in the *K* scale.

These findings accord closely with those reported by other investigators. Gallagher (4) found, in a study of 42 university students treated by the nondirective counseling method, that *D*, *Hs*, *Pt*, and *F* scores decreased significantly following therapy, while *K* increased significantly. Rashkis and Shaskin (8) found significant decreases in *D*, *Hs*, *Hy*, and *Pt* following group therapy. Mosak (7) found decreases in *D*, *Hs*, *Hy*, *Pa*, and *Sc* following individual client-centered counseling in 28 cases. Kaufmann (5) studied 51 university students who had been rated as "improved" following psychiatric treatment, and found significant decreases in *D*, *Hs*, *Pt*, *Sc*, *Pa*, and *F*, and a significant rise in *K*.

At this point a word about the measure of change seems in order. Since the MMPI calls upon the patient to testify concerning himself, and particularly concerning his psychiatric condition, it is subject to all the kinds of distortion to which self-representation is so peculiarly liable. Taken in a clinic, where it is presumably to be scrutinized closely by persons who can be, or already are, very significant in one's life, it is likely to become a communication beyond and apart from the overt meaning of the words. It may be a plea, or a reproach, or a rebuff, or an act of admiration; in brief, an interpersonal communication which expresses something of the relationship of the individual patient to his therapist and to the clinic.

This being inevitably so, it would seem to behoove the psychological researcher to look for evidence of change in other kinds of functions and performances for consideration with the MMPI. In the absence of such other evidence, all that one can do is keep in mind that there is a possibly distorting element of self-representation present in questionnaires. (Perhaps, however, rather too much has been made of the fakeability of tests; life itself can be faked, and many a "cure" is maintained for a time in the interests of an unresolved transference. Furthermore, it may reasonably be argued that durable character traits have a way of showing themselves without one's conscious or even unconscious consent, and that the validity of so well-studied an instrument as the MMPI is not likely to be massively influenced by momentary set.)

Percentage of Improvement

In addition to measuring change simply by considering the average discrepancy score for each scale, it is of some interest to inquire into the *percentage* of change for the better on the scales which do show change. On the scales which most directly reflect neurotic symptoms of which the patient complains (*Hs*, *D*, and *Hy*), the percentage of change in the direction of fewer complaints ranges in the present study from 61 per cent to 67 per cent. This observation concords well with the estimate which Eysenck reports as an over-all rate of improvement with therapy from the data he collated. His judgment was that an estimate of improvement or recovery in two-thirds of the cases would prove "not very far out." The Ego-strength scale, as a matter of fact, showed an increase on second testing in 66 per cent of the cases. None of the other scales changed in the "improved" direction in as many as 60 per cent of the profiles in the sample; the Lie scale, however, did decrease on second testing more than twice as often as it increased (52 per cent vs.

Table 2

Discrepancy Scores ($kt_1 - kt_2$) on MMPI Scales for Treatment Groups and Nontreatment Group

MMPI scale	Individual psychotherapy ($N=42$)			Group psychotherapy ($N=85$)			Nontreatment group ($N=23$)		
	Mean	*SD*	*t*	Mean	*SD*	*t*	Mean	*SD*	*t*
1. Lie	1.79	4.46	2.57*	1.58	5.20	2.78*	3.09	4.84	2.99*
2. *F*	0.74	14.67	0.32	0.32	8.07	0.36	−1.13	4.79	1.11
3. *K*	−2.71	7.48	2.32*	−0.86	6.55	1.20	1.30	5.63	1.08
4. *Hs*	3.73	11.29	2.11*	2.50	9.59	2.39*	3.00	8.01	1.76
5. *D*	8.26	15.63	3.38*	2.58	11.60	2.04*	2.74	9.66	1.33
6. *Hy*	4.43	10.64	2.66*	2.51	8.23	2.79*	3.48	9.37	1.74
7. *Pd*	0.57	12.45	0.29	1.19	9.19	1.19	4.57	9.47	2.26*
8. *Pa*	1.69	10.58	1.02	2.09	6.99	2.74*	1.91	8.25	1.08
9. *Pt*	4.00	14.44	1.77	2.58	11.54	2.05*	1.04	8.56	0.57
10. *Sc*	1.38	16.52	0.53	1.01	10.80	0.86	−2.13	8.61	1.16
11. *Ma*	−0.52	11.26	0.30	0.22	8.71	0.23	−1.61	8.68	0.87
12. *Es*†	−2.62	6.60	2.38*	−1.40	7.14	2.10*	−1.80	3.85	2.04*

* Mean discrepancy significantly different from zero at .05 level of confidence.
† Computations based on raw scores rather than *T*-scores.

25 per cent), but also remained the same much more often than did the other scales.

The immediate general conclusion to be drawn from these data is that the psychotherapy patients, if one is to judge from their MMPI performances, are less ill than when they began treatment. The observed changes are primarily of a symptomatic rather than of a characterological sort: the patients are in a better mood (less depressed), are not as hypochondriacal, and are both less defensive (as shown in the Lie scores) and less given to repression as a defense mechanism (the Hysteria scale). The increase in the *K* score in the individual psychotherapy patients no doubt merits special comment, especially since it was found to increase following therapy in some of the other studies cited. The common interpretation of *K* is that it indicates defensiveness, and that a high *K* score betokens a test-taking attitude which involves representing oneself in a better light than actual clinical condition would warrant. However, in studies over the past five years at the Institute of Personality Assessment and Research, *K* has regularly been observed to correlate positively with favorable aspects of personal functioning. In the present study, it should be noted that while *K* is rising, *L* is falling; if indeed an attempt were being made by these patients to represent themselves in an unwarrantedly better light at second testing, one would expect *L* to rise as well. Perhaps the *K* scale needs a systematic re-evaluation before being accepted as a measure of test-taking attitude.

Changes in the Control Group

Patients who received psychotherapy clearly responded differently to the MMPI when they were tested at the conclusion of the therapy; what, then, of the patients who remained untreated for a comparable length of time? Does time alone have an equal effect, or does treatment make a significant difference? A partial answer is provided by the comparable statistics for the nontreatment group, which are presented in Table 2.

From Table 2 it is evident that the patients who were waiting for treatment had somewhat fewer complaints on second testing, although most of the clinical scales do not change significantly. On the neurotic triad (*D*, *Hs*, *Hy*) the average decrease is slightly greater for the controls than for the group psychotherapy patients. The percentage of change for the better in those scales averaged 61 per cent in the control group, which is slightly, but not significantly, less than in the psychotherapy groups. There is a significant increase in the control group in scores on *Es*, which would suggest that at the end of the waiting period the patients are somewhat more amenable to psychotherapy than at the beginning of that period. The actual percentage of patients who improved on *Es* in the control group was 55 per cent, compared with 66 per cent in the therapy groups, a difference which is not statistically significant.

One might question at this point whether the patients on the waiting list should really be thought of as "untreated." In a sense, of course, simply having committed oneself to participating in psychotherapy, and having had a reciprocal commitment from a clinic to afford psychotherapy, even though not immediately, represents a breaking of the neurotic circle. A force for change has already been introduced. In addition, the initial interview and the psychological testing may themselves be therapeutic events, since during such sessions the patient makes some efforts to confront himself and his problems more objectively than he has in the past. Moreover, the focus of the intake interview at the Kaiser Foundation clinic was generally upon realistic questions concerning the patient's situation and possible courses of action, so that even the waiting-list "controls" might well feel that a realistic review of their situation had been achieved, that an important decision to change had been made by them, and that in general they had obtained more control over their own lives. As the statistics in Table 2 indicate, these changes are slight and are not statistically significant for the clinical scales (with the exception of *Pd*), but there is *some* change. From a methodological standpoint, it may be suggested that other control groups should be added in future studies, since the psychological factors involved in being on a waiting list are themselves possibly determinative of changes in psychiatric condition.

Table 3

Comparison of Treatment and Nontreatment Groups in Terms of Discrepancy Scores ($k_{t_1}-k_{t_2}$) on MMPI Scales

MMPI Scale	Individual psychotherapy vs. nontreatment: Diff. between mean discrepancy scores	Individual psychotherapy vs. nontreatment: *CR*	Group psychotherapy vs. nontreatment: Diff. between mean discrepancy scores	Group psychotherapy vs. nontreatment: *CR*
1. *L*	1.30	1.06	1.51	1.31
2. *F*	1.87	0.75	1.45	1.10
3. *K*	4.01	2.44*	2.16	1.58
4. *Hs*	0.73	0.30	0.50	0.25
5. *D*	5.52	1.75	0.16	0.06
6. *Hy*	0.95	1.49	0.97	0.45
7. *Pd*	4.00	1.41	2.38	1.08
8. *Pa*	0.22	1.25	0.18	0.15
9. *Pt*	2.96	1.11	1.54	0.70
10. *Sc*	3.51	0.68	3.14	1.47
11. *Ma*	1.09	0.44	1.83	0.89
12. *Es*	0.82	0.64	0.40	0.36

* Difference statistically significant at .05 level.

The question remains, is there significantly more change with regular and formally defined psychotherapy than without it? The answer is provided in Table 3, which presents comparisons of the treatment and nontreatment groups in terms of discrepancy scores. When the amount of discrepancy consequent upon nontreatment is subtracted from the amount consequent upon either individual psychotherapy or group psychotherapy, what remains is rarely a discrepancy of such magnitude as to be statistically significant. The only comparison which did yield a significant difference in discrepancy scores was the comparison of the individual psychotherapy group with the nontreatment group on the *K* scale; and, in that instance, the difference was large because in the case of the nontreatment group the discrepancy was positive (decrease in *K* scores), while in the individual psychotherapy group the discrepancy was negative (rise in *K*). For the most part, however, the changes tended to be in the same direction for treatment and nontreatment groups, and of almost equal magnitude.

Summary

The results yielded by this study may be summarized as follows:

1. The mean pre-post MMPI scores of patients undergoing both group and individual therapy showed significant improvement on symptom scales.

2. Patients who decided to enter therapy but who were required to wait for treatment showed some MMPI improvement during the waiting period, primarily in prognosis.

3. The therapy patients did not improve significantly more than did the waiting-list controls.

4. These results suggest that the use of "waiting-list" controls in investigation of changes in therapy may be complicated by therapeutic factors involved in the initial intake evaluations and decisions.

5. The use of additional "nonwaiting-list controls" is suggested for future investigations on the effect of psychotherapy.

Received November 3, 1954.

References

1. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.
2. Barron, F. Ego-strength in normal men as measured by an MMPI scale. Paper read at the West. Psychol. Ass., 1954. Library of the Institute of Personality Assessment and Research, Univer. of California, Berkeley.
3. Eysenck, H. J. The effects of psychotherapy: an evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.
4. Gallagher, J. J. MMPI changes and client-centered therapy. *J. consult. Psychol.*, 1953, **17**, 334–338.
5. Kaufmann, P. Changes in the Minnesota Multiphasic Personality Inventory as a function of psychiatric therapy. *J. consult. Psychol.*, 1950, **14**, 458–464.
6. Leary, T., & Coffey, H. The prediction of interpersonal behavior in group psychotherapy. *Group Psychother.*, 1954, **7**, 7–51.
7. Mosak, H. H. Evaluation of psychotherapy: a study of some current measures. Unpublished doctor's dissertation, Univer. of Chicago, 1950.
8. Rashkis, H. A., & Shaskin, D. A. The effects of group therapy on personality scores. *Amer. J. Orthopsychiat.*, 1946, **16**, 343–349.
9. Rosenzweig, S. A transvaluation of psychotherapy—a reply to Hans Eysenck. *J. abnorm. soc. Psychol.*, 1954, **49**, 298–304.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Brief Psychotherapy and Enuresis[1]

Daniel H. Harris, Richard W. Firestone, and Carl M. Wagner
Marine Corps Recruit Depot, San Diego, California

The present study derives from the premise that nocturnal enuresis in adults is primarily psychogenic in origin. Supporting this was the observation that the psychological climate induced by military training could often be meaningfully related to variations in the enuretic symptoms of Marine recruits. This experiment investigated the effect of a psychotherapeutic situation created prior to training on the adjustment of such recruits. The specific hypothesis tested was: a significant reduction of admissions to the psychiatric observation ward in which enuresis is a factor will result if, during the regular screening procedure, enuretics receive brief psychotherapy combining recognition and acceptance of their problem with specific suggestions for overcoming the habit.

Two hundred recruits (ages 17 to 27) entering training with an admitted history of enuresis during the previous five years were randomly assigned to an experimental group ($N = 100$) or control group ($N = 100$). The two groups did not differ significantly in age, education, intelligence, or degree of apparent neuropsychiatric involvement.

For experimental subjects, the regular psychiatric screening interview was expanded into a brief psychotherapeutic session designed to relieve anxiety related to enuresis by means of a permissive and reassuring discussion of the symptom, plus a focusing of positive motivation around some plausible behavioral suggestions for overcoming the habit. *S*s in the control group had only the regular screening interview. Subsequent contacts with *S*s in either group followed a standardized pattern in which further attempts at therapy were avoided.

On the basis of a rigid set of admission criteria, 20 control *S*s were admitted to the psychiatric observation ward prior to completion of recruit training, while only eight were admitted from the experimental group. Chi-square analysis revealed this difference to be significant, supporting the hypothesis at better than the .02 level of confidence. Apparently the therapeutic situation utilized in this study was instrumental in substantially reducing the number of ward admissions.

Among a third group of recruits who concealed histories of enuresis at screening but later proved to be enuretic, the ward admission rate was five times as high as that for experimental and control groups combined, the difference being significant at the .01 level. This finding gives additional support to the main hypothesis.

Brief Report
Received April 19, 1955.

---

[1] Opinions or conclusions contained in this paper are those of the authors. They are not to be construed as necessarily reflecting the view or endorsement of the Navy Department.

An extended report of this study may be obtained without charge from Daniel H. Harris, Marine Corps Depot, San Diego, California, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4592 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

The Measurement of Depth of Interpretation[1]

Norman I. Harway,[2] Allen T. Dittmann,[3] Harold L. Raush,
Edward S. Bordin, and David Rigler
University of Michigan

When the proponents of one viewpoint of psychotherapy criticize the studies of any other school of thought, the criticisms tend to fall in one or both of two classes: (*a*) the concepts and relationships studied are unimportant or have only minor relevancy; (*b*) the conditions of measurement are such as to preclude proper investigation of the variables under consideration.

The study reported here is one of a series of investigations of psychotherapeutic processes (1). In these studies we have attempted to select variables which have theoretical and practical relevance over a wide variety of therapeutic approaches. Furthermore, we have concerned ourselves intensively with issues of measurement. In these ways our eventual goals of hypothesis testing can, we hope, communicate beyond the narrow barriers of school partisanship.

One of the major activities of the psychotherapist in the course of treating his patient is to engage or avoid engaging in interpretive behavior (5, 9). Collier has presented a rather full discussion of the interpretive continuum as it relates to psychoanalytic and nondirective therapy (2). Like Collier's study (3), the present paper reports the development of a scale for the measurement of a therapist's depth of interpretation. We felt, however, that the immediate utilization of such scales for hypothesis testing would be premature. In studies of psychotherapy, as in all studies involving ratings and codings, interpretation of data is limited by our understanding of the conditions of data collection. So long as these conditions are not understood and accounted for, investigators will arrive at discrepant results, and will question one another's data.

Typical among measurement issues raised with regard to variables in psychotherapy are questions concerning context, unit size, and method of data presentation. Each of these issues becomes the focus for sometimes naive, sometimes sophisticated criticism of published studies. For example, we may consider the kinds of statements that are apt to be made in discussions of research in psychotherapy: "The failure to provide judges with information as to what has gone on previously in therapy, makes their judgments of the behavior in question meaningless" (the problem of context); or, "What is *really* meant by the concept is not any specific, momentary behavior, but rather the over-all effect" (the problem of unit size); or, "The judges would have to have movie films (or recordings) in order to *really* evaluate the behavior" (the problem of mode of presentation).

The specific definition of depth of interpretation used in the present series of studies is: "Any behavior on the part of the therapist that is an expression of his view of the patient's emotions and motivations—either wholly or in part—is considered an interpretation. A patient has varying degrees of awareness of his emotions and motivations. Depth of interpretation is a description of the relationship between the view expressed by the

---

[1] This paper is part of an investigation supported by a research grant M–516 C–2 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The authors are indebted to Dr. R. L. Cutler, Dr. W. L. Hays, Mr. Z. Sperber, and Miss J. V. Williams for their assistance in various aspects of the research.

[2] Now at the University of Rochester, Rochester, New York.

[3] Now at the National Institute of Mental Health, Bethesda 14, Maryland.

Bureau Ed Psy. Research
ING COLLEGE
DAVID
Dated
No

therapist and the patient's awareness. The greater the disparity between the view expressed by the therapist and the patient's own awareness of these emotions and motivations the deeper the interpretation." The definition is compatible with both psychoanalytic and nondirective viewpoints as expressed in the literature (5, 9) and is sufficiently broad to encompass the data of a variety of therapeutic approaches.

Acceptance of the definition of interpretation as an interaction of patient and therapist has certain implications for the measurement operations discussed above:

(*a*) *With regard to unit size:* Since we could see no firm basis for an a priori choice as to unit, we felt that a comparison of extremely divergent sizes would best indicate any effects that might go with variation in size.

(*b*) *With regard to method of presentation:* Since the variable of depth of interpretation seems in the main to deal with the relation between the content of the therapist's behavior and the content of the patient's awareness, we should expect little or no difference in the evaluations obtained by using the typescript of an interview as compared with using the actual sound recording.

(*c*) *With regard to context:* Clearly the interpretive value of a therapist's statements is not determined *in vacuo* but in relation to the determination of what the patient "actually" means and how conscious or aware he is of this actual meaning. Thus, if we force judges to rate depth of interpretation without giving them information on what the patient has said or is saying, the ratings should have little or no correlation with ratings obtained when the necessary information is available.

Development of Rating Scale

Following Thurstone's method of equal-appearing intervals (10), a seven-point scale for rating depth of interpretation was developed. Seventy statements descriptive of therapist activity were sorted by two groups of judges, one experienced, the other naive. The experienced group consisted of fourteen staff members and advanced graduate students in clinical psychology, each with at least 100 hours of experience in psychotherapeutic practice. Eleven undergraduate students in the first course in psychology composed the "naive" judges. Judges with disparate experience were used in hope of identifying items highly communicable to a wide range of possible raters.

The importance of taking judge groups into account was demonstrated by the fact that the naive judges were significantly more variable and placed significantly more items at the deeper end of the scale than did the experienced judges.

Seven of the statements had been originally selected so as not to fit the requirements of our definition; while descriptive of therapist activity they did not give any indication of patient awareness or of the relation of the therapist's activity to the patient's awareness, e.g., "Questioning designed to gain information," "Therapist suggests that the patient desires to terminate therapy." These seven statements offered an internal check. If they scaled along with our bona fide statements, the meaning of the resulting scale would be open to question. None of these "spurious" statements met our criteria for inclusion in the scale.

A seven-point graphic scale was obtained. Nine statements were selected to exemplify points on this scale. The statements and their scale positions (running from low to high depth) are:

(*a*) Therapist merely repeats the material of which the patient is fully aware (1.1).

(*b*) Restatement of material of which the patient is aware (1.8).

(*c*) Implied focusing with regard to material of which the patient is aware (2.3).

(*d*) Therapist connects for the patient two aspects of the content of the previous patient statement (3.4).

(*e*) Reformulation of the behavior of the patient during the interview in a way not explicitly recognized previously by patient (3.9).

(*f*) Therapist comments on patient's bodily and facial expressions as manifestations of patient's feelings (4.4).

(*g*) Use of preceding patient statement to exemplify a process that has been building up during an interview and of which the patient is seemingly unaware (4.9).

(*h*) Therapist speculates as to a possible childhood situation that might relate to current patient feeling (5.9).

(*i*) Therapist response deals with inferences about material completely removed from the patient's awareness (6.9).

In the studies below the definition of depth of interpretation and the scale were presented with the statement that the nine items on the scale "are not inclusive of all possible descriptions of depth at these points . . . but are merely examples of types of things that could occur and where they would fall on the scale."

The Conditions of Measurement Study

Design

Sixteen raters used the scale to evaluate four therapy interviews, each taken from a different case and each involving a different therapist. The raters had the same qualifications as those of the experienced judges involved in the development of the scale. The three sets of measurement conditions discussed earlier were investigated. These were: (*a*) the unit of the interview to be rated—the individual therapist response vs. the interview as a whole, (*b*) method of presentation of the interview to the rater—typescript vs. tape recording, (*c*) amount of information or context given to the rater—the interview to be rated with its preceding interview (preceding interview condition) vs. only the therapist's responses in sequence from the interview to be rated (therapist only condition).

The study was designed to control systematically or evaluate the influence of raters, interviews and practice in the use of the scale. The time-consuming nature of the task did not allow complete control of these factors. A relatively great amount of time is necessary to read or to listen to a number of 50- or 60-minute interviews. Also to be considered were the number of raters available with appropriate experience and training and the number of clearly recorded psychotherapy interviews obtainable.

A 2 × 2 × 2 factorial design was projected (each of the three variables—method of presentation, context, and unit size—was varied over two levels). To form a latin square such that each rater rated under each combination of the experimental conditions in a different order than any other rater, and such that each combination of the experimental conditions was rated but once at each order position from one to eight, required eight raters. Table 1 presents the eight possible combinations of the three factors and the assignment of the raters to the various combinations. The design was replicated, giving us a total of sixteen raters. Kogan (7) has noted the advantages in certain instances of "confounding" in the design of psychological experiments. In the present instance, confounding of the three-factor interaction (Presentation × Context × Unit) with the rater differences reduced the interviews to four and the time required by half.[4]

Table 1

Assignment of Raters to Experimental Conditions

	Typescript		Recording	
Context	Interview unit	Response unit	Interview unit	Response unit
Therapist only	I*	V	V	I
	II	VI	VI	II
	III	VII	VII	III
	IV	VIII	VIII	IV
Preceding interview	V	I	I	V
	VI	II	II	VI
	VII	III	III	VII
	VIII	IV	IV	VIII

* Roman numerals represent raters.

When the interview was the unit rated, only one score was obtained from the rater. When the response was the unit rated, there were, of course, as many scores for each rater as there were responses. In the latter case the response unit ratings for a single judge were averaged to give a rating for the interview as a whole, and then mean response unit ratings were compared with ratings of interviews as units. Regardless of the method by which these ratings were obtained, they are assumed to represent measurements of one aspect (i.e., depth of interpretation) of the same fundamental variate (behavior). Although some of the entries in this analysis were arrived at by averaging a set of ratings of responses, these averages are not to be interpreted in the usual sense of representing a number of independent observations of the fundamental variate. They are, instead, meant to represent only ratings of behavior arrived at in a particular way—the rating of a whole by means of ratings of component parts.

---

[4] Among the characteristics of these interviews: The settings were a University counseling center, a Mental Hygiene Clinic, an out patient clinic and a NP Hospital. The hours represented were the 20th, 40th, 146th and 11th. Therapists were of analytic or modified analytic orientation. They ranged in experience from less than 1 year to 10 years of therapeutic work.

Another problem in connection with the use of response units was the pertinence of depth of interpretation to the various responses in an interview. This was resolved by instructing raters to rate each unit but to indicate those units or responses which they did not feel were ratable on this particular scale. The ratings to responses so indicated were considered as "forced" ratings and were not used in computing the means for the response unit scores nor were they used in computing interrater agreements in rating.

Results

As seen in Table 2, the scale of depth of interpretation significantly differentiated among the interviews at the .01 level. In fact, the variance for interviews was about twice that attributable to pooled experimental conditions. The experimental conditions were found to be a significant source of variance at the .01 level. Any interview by experimental condition interaction, if such existed, could not be evaluated.

Variation due to sequences was not significant. The difference between raters within a sequence was significant at the .05 level. Such a finding could result from the use by raters of different portions of the scale. It need not affect the agreement between raters in the ordering of responses or interviews on the scale.

Unit size. Since each experimental condition represents a specific combination of three

Table 2

Analysis of Variance

Source of variation	Sum of squares		*df*		Mean square		*F*
Sequences	5.69		7		.81		1.89
Raters within sequences	8.50		8		1.06		2.47*
Total between raters	14.19		15				
Order	.18		3		.06		—
Interviews	12.64		3		4.21		9.78**
Experimental conditions	15.23		6		2.54		5.89**
Presentation		.30		1		.30	—
Context		.49		1		.49	1.14
Units		13.32		1		13.32	30.93**
Presentation × context		.42		1		.42	—
Presentation × units		.49		1		.49	1.14
Context × units		.20		1		.20	—
Pooled error	15.51		36		.43		
Greco-Latin square error		4.61		12		.38	
Pooled rater × order interaction, different sequences		10.90		24		.45	
Total within raters	43.56		48				
Grand Total	57.75		63				

* Significant at .05 level.
** Significant at .01 level.

factors, the sum of the squares and degrees of freedom could be broken down into the component parts. Testing each of the component mean squares for the experimental conditions against the pooled error mean square indicates that the difference in ratings due to experimental conditions was primarily the result of one factor, namely, the size of the unit evaluated. For each interview the evaluation based on the interview unit was "deeper" than the mean of the response unit ratings. A further study showed that this result was not a function of the exclusion of forced ratings. Apparently the judge's perception of the depth of an entire interview is strongly determined by the therapist's deeper responses. This finding is of particular significance because previously many investigators have made the assumption that the rating of an entire interview was equivalent to the pooling of a sequence of response ratings within an interview in terms of some average rating. As Mowrer, Hunt, and Kogan note in a slightly different context, "the fact that a client mentions one problem ten times and a second problem just once does not necessarily mean to the clinician that the first problem is ten times as important to the client" (8, p. 259).

Method of presentation. It had been expected that ratings of depth of interpretation would not differ when based on a typescript as compared to a tape recording of a therapy interview. The basis for this expectation was the fact that this variable was primarily a "content" variable, that is, that its rating depended upon the actual words spoken, and not upon some other factor such as tone of voice, facial expression, etc. The analysis in Table 2 indicates that there is no difference in the ratings of interviews under different methods of presentation of the interview.[5] For evaluations of this particular variable, then, we can in the future present typescripts to our raters and enjoy the practical advantages stemming from the easier administration permitted by this method.

Context. We turn now to the critical question of context. The concept of interpretation, and the specific definition used in this study imply that knowledge of both patient and therapist activity is necessary for a correct appraisal of the interpretive level. If our measurements—in this instance ratings—are valid, then the removal of the patient responses should lead to different results from those obtained when the patient responses are included in the information given to the raters. The analysis of variance presented in Table 2 indicates that the expected difference did not occur. Furthermore the reliabilities of the two methods are comparable, and the cross correlations indicate that the methods are equivalent.

As has been noted, such a finding casts doubt on the validity of our measurements or on the adequacy of the definition of the concept. A third alternative, however, suggested itself. It seemed possible that the raters were obtaining information as to the patient's behavior from the sequence of the therapist's responses. To check on this, a study was done in which four new raters and two of the interviews were used.

In this study all ratings were made on the basis of typescript and with response as the unit. Each rater evaluated one interview with only the therapist responses in sequence present (this is the therapist only context of the earlier study) and the other interview with only the therapist responses present but in *random* order. The order of presentation of interviews and of experimental conditions was systematically varied.

There is no difference in the mean response unit ratings between ratings based on sequential therapist only information and ratings based on randomized therapist only information. However, the analysis of interrater reliability coefficients and the cross-context correlations, suggests that marked differences do exist between the sequential context and the random order context. The reliability coefficients for sequential order were .64 and .51, while those for random order were .49 and − .03. On the other hand, none of the eight cross-context correlations differ significantly from zero, ranging from − .19 to .15 with a median of − .05. In other words, while raters might agree within the random order condition, there was no evidence of significant agreement across conditions. These data suggest that the sequence did give the judges enough information about what the patient's behavior may have been to allow them to make ratings equivalent to those of judges who actually had read the intervening patient response. Randomizing the sequence of therapist responses probably randomized the judge's hypotheses about the level of awareness demonstrated by the intervening patient behavior. However, this conclusion was based on a small number of correlations, and one of the two correlations under random order condition was almost as high as those observed under sequential order. Therefore two more judges were asked to rate the therapist responses in this interview under the same two conditions. The four additional sequential order correlations ranged from + .81 to + .52, confirming our previous finding. The four random order correlations were − .03, − .06, − .08 and + .24. The ten additional cross correlations ranged about zero, some, however, reaching the level of statistical significance. These additional data allow us to state with increased confidence that judges were attempting to judge therapist's depth of interpretation as it was defined.

---

[5] Studies of the effects of inclusion of forced ratings, and studies of the reliabilities of each of the methods and the cross correlations between methods lend further support to the conclusion that evaluations of depth of interpretation do not vary as a function of these alternative methods of presentation.

In a further effort to generalize our findings the basic four-judge design was applied to two new interviews. We were surprised to find that the sequential order, random order and cross correlation were all at the same level, median correlations being + .57, + .61, and + .53, respectively. Our present impression is that the therapist style in these interviews (early counseling interviews) will explain the result. Therapists sometimes recapitulate to the patient what he was saying or doing before presenting an interpretive reaction to the patient's activity. Thus, information would be available to a judge no matter where in the sequence the therapist response was located. As a first check on this impression a count was made of the number of words used by therapists in the two studies. As expected, therapists in the interviews used in the latter study tended to make longer responses. The information-bearing characteristics of therapists' communications are now being analyzed and further sequence studies are planned to test these ideas.

Reliability and validity. It might be argued that the failure to find differences in the main study due to context is due to the low reliability of our measure. One of the usual methods for determining reliability of ratings is by means of agreement between raters. Almost all the reliability coefficients based on the ratings of two raters working on the same interview under the same context and presentation conditions were statistically significant. The median coefficient in the main study was .40. Reliabilities of this order are disappointing and at first glance an obvious suggestion would be to dispense with the scale. It must be considered, however, that what is represented by these numerical values is analogous in test development to item reliability. For example, we computed the reliability of response unit ratings of depth for one of our interviews on the basis of a ten-response segment. The means of the ratings for the first ten responses, second ten responses, and so on, were obtained for two raters, and these means were then correlated. Interjudge agreement was now .74 compared to .36 for single response units. As in a test, we may increase our reliabilities by increasing the number of items, each of which may have rather low reliability. The problem, however, remains: Can we study single responses, rather than summing or averaging over responses? Here too, we may make use of test theory. Where one is interested in individual items, where any one judgment may be of low reliability, as in opinion surveys, accuracy of measurement can be attained through increased sampling. We have, for example, found that when we pool the ratings of a number of judges on individual items, we obtain almost perfect correspondence (rhos of .89, .94, and .93) with the ordering arrived at by combined judgments of other judges using other methods, such as rankings and method of similarities by triads (4).[6]

[6] Further confirmation of the usefulness of the scale appears when reliability of individual response items is estimated by the Horst formula (6). Analysis indicates that agreement among the judges accounted for from 39 to 62 per cent of the total variance among items.

Considering the statistical significance of most of the coefficients, the relatively inferential and complex nature of the variable being rated and the fact that none of the raters had any training in the specific use of the scale, the extent of interrater agreement for depth of interpretation is far from disappointing. The certainty with which the four interviews could be discriminated gives substance to this conclusion. Further, if our measurements were sufficiently reliable to differentiate interviews and unit size, it follows that our failure to find differences between the preceding interview and therapist-only context conditions cannot be attributed solely to errors in measurement.

Having considered reliability, we may now summarize our findings concerning the validity of the scale. The later studies indicated that the initial test of validity, based upon a comparison of preceding interview with therapist-only context conditions, was not critical. Further, these later results do provide evidence for validity. Frequently, in the measurement of variables such as the one presented here, validity is not evaluated until evidence is available of success in predicting relationships with other variables. We have illustrated in this paper that it is possible to check on validity of measurement by examining the consistency of the measurements with the definition of the variable.

Summary

In this paper we have illustrated not only how different procedures of measurement lead to differing perceptions of the therapeutic process, but also how varying the measurement conditions can, in certain instances, enable a check on the internal validity of the concept under investigation. Three of the conditions of measurement—method of presentation, unit size, and context—were varied in presenting psychotherapy interviews to raters. Ratings were of the depth of a therapist's interpretations to the patient and were based on a seven-point scale. The four interviews used were clearly discriminated. Variations in unit size and in certain conditions of context led to changes in ratings of depth.

Received October 22, 1954.

References

1. Bordin, E. S., Cutler, R. L., Dittmann, A. T., Harway, N. I., Raush, H. L., & Rigler, D. Measurement problems in process research on psychotherapy. *J. consult. Psychol.,* 1954, **18,** 79–82.
2. Collier, R. M. A basis for integration rather than fragmentation in psychotherapy. *J. consult. Psychol.,* 1950, **14,** 199–205.
3. Collier, R. M. A scale for rating responses of the psychotherapist. *J. consult. Psychol.,* 1953, **17,** 321–326.
4. Coombs, C. H. Theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden Press, 1953. Pp. 471–535.
5. Fenichel, O. *Problems of psychoanalytic technique.* Albany, New York: Psychoanalytic Quarterly, 1941.
6. Horst, P. A generalized expression for the reliability of measures. *Psychometrika,* 1949, **14,** 21–31.
7. Kogan, L. S. Variance designs in psychological research. *Psychol. Bull.,* 1953, **50,** 1–40.
8. Mowrer, O. H., Hunt, J. McV., & Kogan, L. S. Further studies utilizing the discomfort-relief quotient. In O. H. Mowrer (Ed.), *Psychotherapy theory and research.* New York: Ronald, 1953. Pp. 257–295.
9. Rogers, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
10. Thurstone, L. L. Attitudes can be measured. *Amer. J. Sociol.,* 1928, **33,** 529–554.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Responses to Imagined Auditory Stimuli as Compared to Recorded Sounds[1]

David R. Stone
Utah State Agricultural College

The use of auditory stimuli as a source of useful data is in the exploratory stage. Preliminary work has indicated that such an approach can produce useful diagnostic signs, both for clinical and developmental analysis.

The work to date has centered around two major trends, (*a*) presenting sounds in single order, requiring free associative response to the sounds in turn (1), and (*b*) presenting the sounds in sets, asking for multiple association to the sounds taken as a whole (2).

The present study followed the latter trend, and was an amplification of the multiple-association technique as applied to *imagined* sounds, rather than to actual recorded sounds. The sounds were described verbally, in contrast to being played.

In the recorded version, the subjects (220 college students) heard a bird chirping, a train coming from a distance which comes closer until a crashing is heard, and a trumpet fanfare.

The subjects for the study of imagined sounds were 74 students in a general psychology class. The instructions were:

> In a moment I will describe three sounds. When you hear the descriptions, write a story in which you use the sounds as described. Tell what has led up to the sounds, what is happening, and how it turns out. You will have five minutes. For the first sound, imagine that you hear a bird chirping. For the second sound, imagine that you hear a train puffing in the distance. You hear the faint sound of a whistle. Gradually it gets closer and closer. Then you hear a terrific crash. For the third sound, imagine that you hear a trumpet fanfare.

At the end of four minutes, the subjects were told of the time and advised that they would have one minute to complete their story by telling how it turned out.

In terms of length, order of sounds, situations described, characters, sound descriptions, and outcomes, the responses to imagined sounds were either similar to the recorded version, or included them plus additional elaboration. The use of imaginary (verbally described) sounds seems as productive, though less structured than the recorded stimuli.

Perhaps a more subtle analysis would reveal differences not now apparent. It is presumed at present that the association to multiple stimuli is a more crucial element of these responses than whether the auditory stimuli are imagined or real.

Brief Report
Received April 25, 1955.

References

1. Wilmer, H. A., & Husni, May. The use of sounds in a projective test. *J. consult. Psychol.*, 1953, **17**, 377–383.
2. Stone, D. R. *The auditory apperception test.* Beverly Hills, Calif.: Western Psychological Services, 1953.

---

[1] An extended report of this study may be obtained without charge from David R. Stone, Utah State Agricultural College, Logan, Utah, or for a fee from the American Documentation Institute. To obtain it from the latter source, Order Document No. 4591 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm, or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Clinical Statements as Scientific Propositions and Social Decisions[1,2]

Frank M. du Mas
Montana State University

"Tom is a neurotic." "Dick is to remain at home under the continuing care of his parents until further study of his case can be made." "Harry is to be committed to the state institution for mental defectives." These statements, though overly simplified and out of context, are the kinds of statements clinical psychologists must make in their routine professional work. Each statement is different from every other statement in certain respects. The statement about Tom is a more or less straightforward statement of a matter of fact. The statement about Dick is a social decision to withhold positive action. The statement about Harry is a social decision to initiate a definite course of action. The clinical psychologist is an applied scientist and these statements are relevant to the *applied* part of his work. He is a scientist operating within the confines of his social matrix and he is constrained to obey administrative protocol, the laws of his society, and to make decisions about a patient which may radically change the patient's way of life. It is not a question of whether or not a clinician wants to make these kinds of decisions. He must, if he is to work effectively in a social institution such as a hospital.

But the clinician is also a scientist. In his endeavor to push back the frontiers of knowledge, he frequently resorts to theoretical analysis in his study and construction of theoretical systems which organize and integrate his field. When he does this, he turns to deductive logic in order to facilitate his theoretical work. He often restates an inductive generalization as a postulate in a deductive system. Then, as in all theoretical work, he operates within a formal system and the rules of the game are not quite the same as the rules appropriate to inductive empirical science or the rules of society.

It seems, therefore, that the clinician is concerned with statements in three domains: formal, empirical, societal. What is the appropriate logical frame of reference to use in evaluating the validity of statements in each of these domains? Since, as we shall show, eminent authorities give quite different answers to this question, let us refine the issues.

One widely held point of view is well stated by Ayer:

> Like Hume, I divide all genuine propositions into two classes: those which, in his terminology, concern "relations of ideas," and those which concern "matters of fact." The former class comprises the *a priori* propositions of logic and pure mathematics, and these I allow to be necessary and certain. . . . That is, I maintain that the reason why these propositions cannot be confuted in experience is that they do not make any assertion about the empirical world, but simply record our determination to use symbols in a certain fashion. Propositions concerning matters of fact, on the other hand, I hold to be hypotheses, which can be probable but never certain (3, p. 31).

Stated differently, what Ayer is saying is that *formal* propositions may be evaluated by reference to a 2-value logic, that is, any formal proposition is either true or false. In probability terms, the probability of any for-

---

[1] This article is a major revision of a paper read at the 1954 Convention of the Midwestern Psychological Association.

[2] The ideas presented here are relevant to all applied sciences. The discussion is pointed at one particular field, clinical psychology, in order to simplify the discussion. An understanding of our ideas in one applied science should easily transfer to other applied sciences.

mal proposition being true is either 0 or 1. *Empirical* propositions, however, can never be certainly true or certainly false—they can only be probable. That is, an empirical proposition has a truth value which may be represented by a point in a truth continuum. In probability terms, the probability of an empirical proposition being true is always greater than 0 but less than 1. The truth (or probability) continuum implies that a multivalue logic is necessary for the evaluation of empirical propositions because a continuum contains an infinite number of possible points or truth values.

Another point of view maintains that it is possible to evaluate a proposition about an object, person, or event as either certainly true or certainly false. This point of view is implicit in the writings of Allport:

> The truth of the matter seems to be that *this* delinquent has either 100 per cent certainty of becoming a repeater or 100 per cent certainty of going straight (2, p. 156).

In another source Allport states:

> Indeed, if we believe in determinism at all, his (the parolee) chances are either zero or 100: he is bound to succeed or else to violate . . . and (ideography) will bring your prediction . . . to a 100% certain individual prediction (1, p. 17).

Others agree with Allport. Snygg and Combs state:

> Far from being inconsistent, behavior becomes characterized by complete predictability (4, p. 58).

Snygg and Combs (4, p. 437) affirm their agreement with Allport and include a quotation from his work.

Stated differently, what these writers are saying is that *empirical* propositions may be evaluated by reference to a 2-value logic, that is, that it is at least sometimes possible to say that an empirical proposition is certainly true or certainly false. In terms of probability, they say the probability of some empirical propositions is either 0 or 1.

Table 1

Logical Frames of Reference and Their Applications to Specific Clinical Statements

Relations	Truth values*	Applications
I	$0 \leq p \leq 1$	All clinical statements
II	(F) (T)	Formal propositions
III	$p = 0, 1$	Formal propositions
IV	$0 < p < 1$	Empirical propositions
V	(F) (D) (T)	Subjective social decision: withholding action
VI	$0 < F \vert D \vert T < 1$	Objective social decision: withholding action
VII	(F) (T)	Subjective social decision: action
VIII	$0 < F \vert T < 1$	Objective social decision: action

* Parentheses represent discrete truth values; vertical lines, cuts in the truth or probability continuum.

These representative quotations raise certain questions which need to be answered. What logical frame of reference is appropriate to formal statements? To empirical statements? To societal statements? How many and what kinds of logical frames of reference are required by the clinician? Can a unitary, systematic integration be made of the several logical frames of reference?

Let us attempt to answer these questions.

Analysis and Development

The most general model of our ideas is shown in Table 1, relation I. It states that for any proposition the probability, p, of it being true is some value equal to or greater than zero and equal to or less than one. Classical logic holds, however, that there are two truth values, i.e., false, F, or true, T; as shown in Table 1, relation II. We know now that relation II is relevant to formal propositions only. But relation I implicitly contains relation II since by simple translation we get relation III. That is, the *limits* of the truth or probability continuum suffice for the truth values of classical Aristotelian logic.

But no empirical proposition is ever certain, that is, p can never be 0 or 1 for such propositions because (a) there always seems to be error in human observations, (b) no empirical proposition is ever regarded as absolutely true or absolutely false by all those competent to judge, i.e., by all scientists at all times, and (c) all possible observations never have been and never will be made regarding any empirical matter. Item c seems to be true because (i) positive or negative instances may occur in the future and (ii) even in a specified time interval in which observations are made, all possible observations are not made nor do all scientists at that time make observations of the specified phenomenon; therefore, negative or positive instances could occur in the present, could have occurred in the past, and could occur in the future. Hence, that part of relation I which is appropriate to empirical propositions is given by relation IV.

In the practical world of affairs, instances constantly arise about which the clinician wishes to postpone judgment or opinion until he has obtained more information, made more observations, consulted with others, and weighed the evidence before he makes a diagnosis, decides on appropriate therapy, or initiates a course of action. When such instances arise, the clinician often makes a subjective analysis of a patient's case history, test results, present behavior, etc. For example, in the case of Dick the clinician's judgment was to postpone or withhold a definite course of action. The empirical proposition on which his decision rests might be something like "Dick is psychotic." At the time a *subjective* social decision had to be made, the proposition "Dick is psychotic" remained in doubt. He requires, therefore, a 3-value logic: false, doubtful, and true. Relation V is the appropriate logical frame of reference in this instance. However, when an acceptable estimate of p can be calculated, three intervals of relation IV can suffice for the three discrete truth values of relation V. The calculated estimate of p would permit an *objective* social decision to be made that definite action should be withheld in Dick's case if p fell within the D-interval of relation VI.

It seems obvious that a 3-value logic is a cautious guide to decision making in a practical world of affairs. It permits us to postpone a decision until we are surer of the factors involved and their relative weights. But sooner or later the clinician is forced by the demands of the society in which he lives and works to make another kind of decision. That is, societal demands often force him to say "yes" or "no"; this will work or that will not work. He must decide whether a person was or was not morally responsible at the time of a crime, whether a patient is or is not a schizophrenic, whether a person is or is not to be committed to a mental hospital, whether a person will or will not violate his parole, whether a patient is or is not to be released from the hospital, etc. The clinician must make many decisions of an "either-or" nature. The statement regarding Harry is of this sort.

At first glance it may seem that a 3-value logic also would be appropriate to a social decision initiating a definite course of action. But there are serious theoretical as well as practical limitations involved. Situations often arise where "close" decisions have to be made. The D-category of relation V or the D-interval of relation VI may be so wide that a proposition must remain forever in doubt. His files would become full of unresolved cases. Before long, social pressures (administrative, parental, legal) would be brought to bear on him to make definite decisions regarding these cases.

The empirical propositions upon which the statement about Harry rests is "Harry is a mental defective." The social rule, say, is that all those classified as mentally defective must be committed to the state institution for mental defectives. The clinician may study at great length all available information regarding Harry and then make the *subjective* social decision that Harry is a mental defective and, therefore, must be committed to the state institution. Relation VII is applicable here. However, when the clinician can calculate an estimate of p, he may define two intervals of relation IV and relation VIII results. Relation VIII is useful in making *objective* social decisions which initiate a definite course of action.

The preceding development indicates that of the many possible frames of reference (1-, 2-, 3-, . . . , n-value), there are three which have particular relevance to clinical statements: 2-value or 2-interval, 3-value or 3-interval, and multivalue. The concept that unifies them all is the truth or probability continuum.

Summary

The clinical psychologist must serve two masters: science and society. His professional statements will fall in three domains: formal, empirical, and societal. Statements with societal significance are often implictily based upon empirical propositions. Each formal or empirical proposition can be regarded as having a certain probability, p, of being true. The general model is the truth or probability continuum: $0 \leqq p \leqq 1$. Clinical statements about a patient are directly or indirectly empirical propositions.

A clinician has two kinds of social decisions to make about a patient: to withhold a course of action or to initiate a definite course of action. His *subjective* social decisions require a 3-value logic (false, doubtful, true), whenever he wishes to withhold action. His *subjective* social decisions require a 2-value logic (false, true) when he must initiate a definite course of action. He can, however, make *objective* social decisions whenever an estimate of p can be calculated. The 3-value and 2-value logics may then be regarded as 3-intervals and 2-intervals respectively of the probability continuum.

Received December 8, 1954.

References

1. Allport, G. W. The psychologist's frame of reference. *Psychol. Bull.*, 1940, **37**, 1–28.
2. Allport, G. W. *The use of personal documents in psychological science.* New York: Soc. Sci. Res. Counc., Bull., 1942, No. 49.
3. Ayer, A. J. *Language, truth and logic.* London: Victor Gollancz, 1949.
4. Snygg, D., & Combs, A. W. *Individual behavior.* New York: Harper, 1949

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Delayed Gratification and Ego Development: Implications for Clinical and Experimental Research

Jerome L. Singer
Franklin D. Roosevelt VA Hospital, Montrose, New York

Since its initial delineation in Freud's *The Interpretation of Dreams* (13), the concept of delayed discharge has played a pivotal role in the psychoanalytic theory of ego development and the psychology of thinking. Little attempt has been made, however, to explore some of the implications of this concept and to deduce testable hypotheses or to survey systematically some of its parameters in relation to views held by other theorists in psychology. The last decade has witnessed a marked increase of interest in ego functions and a consequent coalescence of psychoanalytic theory and general psychology. It now seems possible to attempt at least an initial survey of the scientific usefulness and validity of Freud's view of delay. In the present paper the concept of delay will be related to the views of other theorists, relevant empirical evidence will be cited, and some implications both for further experimentation and practical utility will be suggested.

Origins in Psychoanalytic Theory

The distinction between *primary* and *secondary* processes in mental functioning has been described by Jones (22) as one of Freud's most profound formulations. The shift from the chaotic, associative, diffuse, and "syncretic" character of primary-process thinking to the more synthetic, goal-directed, reality-oriented secondary process involves as a key step the enforced delay or deferment of gratification (13). As Freud described the sequence of thought development, the gradual internalization of the delaying mechanism provided the clue to the emergence of the reality principle (11). The arousal of a need-tension pressing for discharge results in the infant in a restless discharge of motor activity which eventually subsides upon contact with the satisfying object, e.g., bottle or breast. When such restless motor activity has been repeated frequently in association with the satisfying object, the absence in reality of the object will lead to its hallucinatory representation as soon as the motor innervation is initiated by the aroused need. This step is part of the primary process and reflects the operation of the pure pleasure principle. It can be seen clearly in evidence in the autistic perceptions of psychotics, wish-fulfilling dreams, and hallucinatory food responses of starving castaways. Cathectic energies are still mobile and pressing for immediate discharge at this stage; hence, the ease with which things can stand for other things associated with them only by spatial or temporal contiguity. Chaplin's movie, *The Gold Rush,* portrays this most acutely and humorously. The little tramp with his jerky, restless movements is suddenly transformed into a juicy fowl in the eyes of his starving companion in the snowbound cabin.

Freud at no time brought together all the steps involved in the shift from the hallucinatory representation of the absent drive object to the internalized control of discharge through thought process. Nevertheless, the sequence roughly gleaned from a number of different papers seems to involve a gradual "neutralization" of libidinal energy through identification and incorporation of parental figures, transformation of mobile energies into "bound cathexes," and formation of a functional substructure of the personality, the *ego,* whose properties include motor control, purposive action, directed thought, reality-cognition, and various defensive measures which organize or segmentalize the diffuse innervations arising from basic drives (id) (9, 10, 11, 12, 13).[1] As early as 1900 Freud had already suggested that the transition from mobile to bound cathexes made possible the use of small, controlled energy expenditures at the behest of the secondary process which could serve as signals (13), or as he put it later, "experimental actions" (11). In effect the secondary process substitutes a "thought identity" for the hallucinatory wish fulfillment. Freud wrote, "All thinking is no more than a circuitous path from the memory of a satisfaction (a memory which has been adopted as a purposive idea) to an identical cathexis of the same memory which it is hoped to attain once more through an intermediate stage of motor experiences" (13, p. 602).

Subsequent psychoanalytic theorists have not altered Freud's concepts to any great degree. A brief but stimulating paper by Ferenczi (7) carries Freud's concepts to the clinical sphere and points to some individual differences in motor activity which are coordinated with differences in thought production. Ferenczi cites an example of a patient whose constant foot wiggling would stop whenever she began to think consciously. It became possible for the analyst to use this cessation of motor activity as an indication of resistance if the patient failed to communicate verbally the thoughts that had followed upon her decreased motor response. Fenichel (6), too, has elaborated on Freud's conceptions without, however, altering their sequence or structure to any significant degree, while Anna Freud (8) in her work has emphasized what Hartmann (16) would call the "conflictual ego-sphere" and deals with fantasy (denial) or hypercathected thought processes (intellectualization) only in their defensive role.

The recent work of Hartmann (16), Federn (5), Weiss (56), Hartmann, Kris, and Lowenstein (17), Kris (24), and Rapaport (39, 40) has added new breadth to Freud's conception of the ego. Rapaport (39), in particular, has taken an extremely significant step in the direction of providing a systematic, scientifically verifiable formulation of Freud's concept of delayed discharge. He traces the meaning of this concept at different vantage points: *genetic* (primary to secondary process), *topographic* (unconscious or preconscious and conscious), *structural* (ego functions), *economic* (mobile to bound cathexes or delibidinized energies), and *biological* (adaptive nature of delay and thought). By this analysis it becomes possible to see more clearly the relationship of the psychoanalytic theory of thinking to the work of others less oriented toward concern with the "neutralization" of instinctual energies but more with the adaptive function of thought in biological survival and interpersonal relationships. Rapaport writes, "Thinking explores the possible pathways of action to find the one of least resistance, least danger, and greatest directness, while preserving almost intact the energy necessary for motor action" (39, p. 169). This formulation, in essence unchanged from Freud's formulations in 1900 and 1911, is nevertheless an advance in clarifying the concept of delay without the confusing terminology rooted in Freud's vaguely defined cathectic dynamics. Rapaport's huge volume (40) contains a challenging presentation of the hierarchical, almost functionally autonomous structures of motives and thought processes in the personality structure.

Other Viewpoints of Delay and Thought

The belief that conscious thought or fantasy bears some relation to the inhibition of an overt motor response directed toward some sort of gratification has appeared in the writings of many investigators who worked more or less independently of psychoanalytic theory. The early American functionalists and behaviorists held similar views, even if we except the somewhat extreme and experimentally unsupported approach that thought was *nothing more* than implicit movement of the speech mechanism. James (21) saw "verbal imagination" as the outcome of "suppressed articulation" while Dewey (4) claimed that thinking becomes intensified upon the interruption of a usual sequence of motor acts. Washburn (55) held a somewhat similar point of view, noting also that consciousness and "associative activity" diminish with the overt performance of the act. E. B. Holt, an early pioneer in the integration of Freudian theory with the emerging American behaviorism, wrote in 1915:

---

[1] The meaning of the concept of free and bound cathexes, attributed by Freud to Breuer, is never completely clear in Freud's writings. Breuer (3) seems to have linked "bound cathexes" to "tonic excitation of nervous tension," a phrase of interest chiefly because of its at least superficial similarity to Werner's "sensory-tonic theory of perception" (58).

> Thought is . . . more than the thing thought about: there is active thinking about objects. If we look once more at the least manifold in which cognition occurs—a living organism in, and responsive to, an environment—we see that this further element is the active play of motor attitude, which eventually resolves itself into the less labile but more forceful phenomenon, conduct. Thus, thought is latent course of action with regard to the environment . . . , or a procession of such attitudes (19, page 98).

The theories of the behaviorist and functional psychologists, while less comprehensive in their implications than the psychoanalytic theory, have led at least to some experimental research. These investigations, as summarized by Vinacke (54), have indicated that thought is indeed accompanied by some implicit muscular activity albeit by no means has it been shown to be clearly the *result* of such activity.

More directly related to the psychoanalytic theory, both in influence and implications for a personality theory, are views held by important workers in developmental psychology like Piaget (36), Werner (60), Lewin (29), Murphy (35), and Luria (30). All of these investigators present empirical data to support a view that thought involves the gradual interiorization of motor response or speech which must be inhibited from overt expression because of social restraints or pressures. Luria (30), for example, describes the change with increasing age of children from the vigorous acting out of a battle scene to its representation in drawing with minimal accompanying speech or motor activity. Piaget (36) has traced the change in children's language from the egocentric comments in which everything that comes to mind is pronounced aloud without regard for its implications. Socialization enforces inhibition of a great deal of this speech pattern with the consequent heightening of thought activity, conscious and unconscious. Wolff and Precker (64) describe the sequence of development in terms of the changing reaction of the growing child in response to being struck: at one stage he responds with a direct counterblow, later he may verbally abuse the assailant, and finally he may completely inhibit or delay overt response but fantasize various countermeasures.

Perhaps most precise in its derivation and fruitful in terms of experimental research relevant to the Freudian concept has been the work of Werner (58) and Werner and Wapner (61, 62, 63). Building in part on earlier physiological investigations as well as upon Werner's (60) extensive survey of developmental patterns, these writers have formulated a sensory-tonic theory of perception which offers at least the possibility of a more precise ultimate formulation of Freud's view of delayed discharge in operational, experimentally verifiable terms. Werner and Wapner describe perception as a function of the dynamic interaction of externally derived sensations and pre-existing tonic states in the organism. Both sensory and tonic events have equivalent functions in the economy of the organism and may serve vicariously as substitutes for each other. When available energy is not released in bodily movement the tonicity will express itself in perceptual displacement, e.g., illusory motion, or, perhaps one might add, "movements in the mind's eye," fantasies of motion. There has been an increasing body of experimental literature in support of such a theory of the "vicarious channelization" (62, 63). As an example, the finding reported by Goldman (15) may be cited. Subjects responded to the autokinetic phenomenon under *immobilization, control,* and *heightened motor activity* conditions. Readiness to perceive motion was greatest under immobilization, less under a control condition, and least during activity. The theory and experiments of Werner and his colleagues thus offer a challenging opportunity, but their demonstration of blocked motor discharge leading to perception of movement must still be related more clearly to heightened imaginativeness if the Freudian concepts are to be assimilated into their structure.

Experimental Research with Rorschach's Movement Responses

Perhaps the most striking possibility for operational expression and experimental study of the concept of delay has come in research with the Human Movement response of the Rorschach test. Rorschach (41) in his studies of responses to inkblots delineated a type of response which involved perception of humans in action (*M*). He dealt at great length with his observations that persons who saw several such *M* responses on a blot series tended, on the one hand, to be relatively stable or inhibited in motor activity and, on the other hand, rather imaginative, given to considerable inner living and creativeness or planfulness. In effect, Rorschach, although lacking a clear-cut theory, was formulating a triadic linkage of motor inhibition, perception of motion, and imagination or inner life. Here then appears a possible bridge between the psychoanalytic concept of delayed gratification (inhibition of motor response) and the sensory-tonic theory of vicarious functioning of motor activity and motion perception. The relationship has indeed been emphasized by Rapaport's (40) concern with the delaying capacity of *M*-type individuals and by Werner's (59) work comparing Rorschachs of hyperactive and motorically-phlegmatic feebleminded children.

The heuristic value of Rorschach's *M* response lies in two separate if ultimately related directions. On the one hand we may attempt to test both Werner's and Freud's hypotheses fairly directly by observing what effects emerge when motion is inhibited prior to or during a Rorschach examination. If this delayed discharge of motor activity has the implications described above then, an increase in *M* responses should be a consequence of this inhibition. On the other hand, on the assumption of individual differences in ego organization and control, it is likely that longstanding patterns of motor inhibition or hyperactivity will be reflected in differential Rorschach patterns. Thus, studies of subjects who contrast sharply in Rorschach *M* or in imaginativeness or motor control may throw valuable light on the nature of the delaying mechanism. The first experimental approach is chiefly of theoretical interest and ultimately may lead to the crucial experimentation upon whose results the theory of delay and thought may stand or fall. The second hypothesis, while less crucial theoretically, has important practical implications including the possible development of measuring devices for the vital ego function of delay of gratification and effective use of planning and imagination.

Turning first to the direct consequences of motor inhibition or delay, two experimental studies have reported increases in Rorschach movement responses of normals following periods of motor inhibition, i.e., standing rigidly in awkward postures or voluntary slow writing (32, 48). Another study has investigated the effects of speeding up or slowing down tempos of Rorschach response and has yielded the conclusion that production of *M* responses is indeed related to tendencies or capacities to delay response (45). Further work remains to be done but it seems clear that so far the linkage of delay of motor discharge and motion perception and imagination can be demonstrated directly in normal adults. No work with children has yet been carried out in this direction. A number of studies do suggest that the *M* response emerges most clearly in age groups that are being subjected to the first striking enforced socialization and motor restraint of school entrance (28, 58). It can be said that Freud's early formulation of thought as arising from delayed gratification and discharge of minute quantities of energy in experimental action is certainly consistent with these findings. Whether the subtler refinements of the theory such as the transformation from mobile to bound cathexes or the desexualization of energy leading to hypercathecting of ideational representations can be expressed in operational terms and tested experimentally remains to be seen. Holt (20) has suggested some possibilities in this connection.

The second general avenue of exploration in this field has been more thoroughly traversed because it calls for correlational procedures rather than direct experimental manipulation. Can it be shown that individuals who are given to or capable of extensive impulse control, deferment of gratification or motor control (apart, that is, from athletic skill), will reflect correspondingly both low thresholds for perception of motion and greater imaginativeness? In effect this implies that some people do manifest a persistent style of coming to terms with their environment which involves relative delay of immediate reactions to needs or provocative stimulation and employment of planfulness or fantasy in their mode of relating. Rorschach's distinction between the introversive and extratensive experience-types is clearly an example of such a cognitive style and the testimony of various experienced Rorschach authorities certainly confirms this dimension of personality variation (2, 23, 38, 42).

Actual systematic investigations have linked characteristic modes of motor reactivity to Rorschach movement responses (1, 59). It has also been observed that both normal and schizophrenic subjects who yield numerous *M* responses are more capable of inhibiting motor responses on slow-writing tasks or are less active and responsive to environmental stimuli while observed in a waiting room (32, 47, 49). The other aspect of this relationship, *M* and creativity or imaginativeness, cannot be studied directly. Measures such as the ability to transcend objective stimulus material in thematic apperception stories which may be considered reasonable approximations of this fantasy capacity, have, however, uniformly revealed significant correlations between *M* and imaginativeness, even with general intelligence controlled (1, 18, 37, 43, 44, 47, 53). Recently Singer and Herman (47) reported data revealing that schizophrenics differing in *M* responses differed correspondingly in a series of situations tapping motor control, imaginativeness, spontaneous motor activity, and impulsivity. The results suggested that a common factor does indeed link motility, motion perception, and thought as characteristic ego functions. Motor inhibition has also been related to cognitive inhibition, the tendency to delay verbalization of an old association and to rapidly substitute a new one (31). Finally, some recent work has shown Rorschach's *M* to be related to high levels of aspiration, planning ability, and impulse control (51).

It should be noted, however, that none of the linkages reported in this section represent extremely high correlations. As a matter of fact, considerable variability is present and it is certain that numerous factors ranging from influences of other spheres of ego functioning (16) and from direct situational factors in the face-to-face subject-examiner relationship modify the emergence of the cognitive style in question. It may be concluded, nevertheless, that in general the capacity or tendency to delay response or inhibit motor discharge appears to be a characteristic adjustment of many persons and it is linked in them with tendencies to be sensitive to movements on ambiguous forms or to engage in fantasy activities with relative ease.

Some Genetic Considerations and Clinical Implications

Accepting at least temporarily the last conclusion, what can be said of the way in which these characteristic modes of coming to terms with the environment originate? It seems likely that many temperamental characteristics, particularly gross features such as relative motility tendencies (in which large differences emerge in children [57]) may be constitutionally based. Certainly Freud (12) came to feel that quantitative factors in libido distribution were of such a nature. Nevertheless, one is tempted to seek for considerable additional sources of individual variability in acquired characteristics, modes of adjustment, defensive techniques, or values learned from significant interpersonal situations in the family constellation. One may well ask, what sort of relationships in the family setting foster in the child a capacity to defer gratification or to temporarily restrain motor discharge in response to a need with subsequent resort to the realm of imagination or trial action? A clue is forthcoming in Freud's stress on identification as a means of neutralizing libido and establishing ego defenses and countercathexes. In family constellations where relatively strict patterns of eating, toileting, washing, napping, and recreation are the rule children may more readily, in order to master their tensions, learn to accept delay since the sequence or pattern is consistent. In thus delaying and striving to imitate the adults they find that resort to the level of fantasy gratification is possible and is frequently rewarded by parental figures or leads to new possibilities for mastering the reality situation. Thought and delay become valued (cathected) and organized and a new hierarchical structure of personality emerges, Lewin's level of irreality (29). Within reasonable limits this pattern is highly adaptive and, indeed, one of the most uniquely human characteristics for it frees man from dependence upon immediate stimulation. Yet it can be carried to excess, as in the case where the extreme anxiety of parents leads to overly rigid patterns with the result that unceasing resort to inhibition and delay leads the child into obsessional rumination that gradually loses its reality orientation and creativity.

Lair (26) has presented a review of considerable theoretical material and some data which support the view that an opportunity for identification with benign and loving adults in very early childhood is essential to the development of capacities for motor inhibition and planful fantasy. Goldfarb's studies of adolescents whose first three years of life had been spent in institutions where close relationships with adult figures were not possible are highly suggestive in this connection. These institutional children showed almost uniformly a thinking disorder characterized by lack of time perspective, hyperactivity, absence of "normal inhibitory pattern," and lack of planfulness (14). Kupper (25) has traced in detail how certain types of family constellation lead to development of strong intellectual values and inhibited motor patterns. The work of Mowrer (34) on language development has represented perhaps the first systematic attempt to develop more precisely the manner in which symbolic behavior is learned through attempted identification with the secondary reinforcing characteristics of benign parental figures. Mowrer's formulations in the framework of two-factor learning theory certainly afford the possibility of expressing some of the analytic concepts of identification and "hallucinatory representation" in operational terms and may open this area to experimental study.

To some extent, it would appear, then, that the varieties and opportunities for parental identification may play a role in the capacity to develop delaying mechanisms and suitable imaginative processes. To the extent that social class factors play a role in the organization of family constellations these factors may markedly influence the relative resort to deferred gratification patterns. The middle class family with its routines, family unity at mealtimes, and general emphasis on saving or education provides the child with a setting that fosters such delay. This is particularly the case since the consistency of the pattern leads to the relative assurance that satisfaction will come eventually. In the lower class family parents are less available (particularly fathers), meals are haphazard with no assurance of their being forthcoming, and the value of delay is less clear to the child. Hence we may anticipate that lower class children will show different patterns of means-end cognizance, time evaluation, and imaginativeness. Some tendencies of this sort have indeed been reported in the recent literature (27, 35, 46).

Returning to the Rorschach movement response as a possible measure of delaying capacity and fantasy we might expect that high-M subjects would show more positive feelings toward parents or at least greater identification with them. Studies with the TAT have offered some supporting results but considerably more work remains to be done here (44, 50). It would be important to study known family constellations and observe the patterns of delay, motor inhibition, or fantasy which emerge in the children. One possibility might be exploration of cultural groups known to differ markedly in family structures, e.g., Irish and Italian lower class.

In conclusion, then, it seems plain that Freud's original concept of delayed gratification leading to thought as experimental action is indeed a fruitful one. It is supported in rough outline at least by the observations of a host of investigators and by considerable experimental research particularly that with the Rorschach M response. The dimension of motor inhibition—motion perception—fantasy seems a strikingly useful one, capable of empirical study and providing an interesting approach to the study of personality structure in the family setting. The concept of delay or impulse control as a basic ego function affords also a highly significant contact point between psychoanalytic theory and general psychology. More than fifty years after its original pronouncement, Freud's theory of thought development is vigorously alive and beckons toward new vistas of research.

Summary

This paper represents an initial attempt to place Freud's concept of delayed gratification and thought development in the perspective of a broader framework of theory and research in psychology. After tracing the development of this concept in psychoanalytic literature, points of contact with early American behaviorist and functional psychological points of view were noted. Experimental research that bore on the theory was cited with particular emphasis on Werner's sensory-tonic theory of perception and the empirical data which has grown out of Rorschach's linkage of motor inhibition, motion perception, and imagination. Some implications of this concept for the study of identification processes, internal control mechanisms, and differential parental influences in the family constellation were suggested.

Received November 5, 1954.

References

1. Barron, F. Threshold for perception of human movement in inkblots and its personality correlates. *Amer. Psychologist,* 1953, **8**, 317–318. (Abstract)
2. Beck, S. J. *Rorschach's test.* Vol. II. New York: Grune & Stratton, 1946.
3. Breuer, J., & Freud, S. Studies in hysteria. In *The standard edition of the complete psychological works of Freud.* Vol. 2. London: Hogarth Press, 1952.
4. Dewey, J. *How we think.* New York: Heath, 1933.
5. Federn, P. *Ego psychology and the psychoses.* New York: Basic Books, 1952.
6. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
7. Ferenczi, S. Thinking and muscle innervation. In *Further contributions to the theory and technique of psychoanalysis.* New York: Basic Books, 1952. Pp. 230–232.
8. Freud, Anna. *The ego and the mechanisms of defense.* New York: International Universities Press, 1946.
9. Freud, S. *The ego and the id.* London: Hogarth Press, 1923.
10. Freud, S. *The problem of anxiety.* New York: Norton, 1936.
11. Freud S. Formulations regarding the two principles in mental functioning. In *Collected papers.* Vol. IV. London: Hogarth Press, 1946. Pp. 13–21.
12. Freud, S. *An outline of psychoanalysis.* New York: Norton, 1949.
13. Freud, S. The interpretation of dreams. In *The standard edition of the complete psychological works.* Vol. V. London: Hogarth Press, 1953. Pp. 339–628.
14. Goldfarb, W. Psychological privation in infancy and subsequent adjustment. *Amer. J. Orthopsychiat.,* 1945, **15**, 249–254.
15. Goldman, A. Studies in vicariousness: degree of motor activity and the autokinetic phenomenon. *Amer. J. Psychol.,* 1953, **66**, 613–617.
16. Hartmann, H. Comments on the psychoanalytic theory of the ego. In *The psychoanalytic study of the child.* Vol. V. New York: International Universities Press, 1950. Pp. 74–96.
17. Hartmann, H., Kris, E., & Lowenstein, R. Comment on the formation of psychic structure. In *The psychoanalytic study of the child.* Vol. II. New York: International Universities Press, 1947. Pp. 11–38.
18. Hays, W., Gellerman, S., & Sloan, W. A study of the relation between the verb-adjective quotient and the Rorschach experience-balance. *J. clin. Psychol.,* 1951, **7**, 224–227.
19. Holt, E. B. *The Freudian wish and its place in ethics.* New York: Holt, 1915.
20. Holt, R. Implications of some contemporary personality theories for Rorschach rationale. In B. Klopfer, Mary D. Ainsworth, W. G. Klopfer, & R. Holt (Eds.), *Developments in the Rorschach technique.* Vol. I. Yonkers, N. Y.: World Book, 1954.
21. James, W. *The principles of psychology.* New York: Holt, 1890.
22. Jones, E. *Life and work of Sigmund Freud.* Vol. I. New York: Basic Books, 1953.
23. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book, 1942.
24. Kris, E. On preconscious mental processes. *Psychoanal. Quart.,* 1950, **19**, 540–560.
25. Kupper, H. L. Psychodynamics of the "intellectual." *Int. J. Psychoanal.,* 1950, **31**, 85–94.
26. Lair, W. S. The psychoanalytic theory of identification. Unpublished doctor's dissertation, Harvard Univer., 1949.
27. Levi, J., & Kraemer, Doris. Significance of a preponderance of human movement responses in children below age ten. *J. proj. Tech.,* 1952, **16**, 361–365.
29. Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

30. Luria, A. P. *The nature of human conflicts*. New York: Liveright, 1932.
31. Meltzoff, J., & Levine, M. Motor and cognitive inhibition. *J. consult. Psychol.*, 1954, **18**, 355–358.
32. Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.*, 1953, **21**, 400–410.
33. Mitchell, H. E. Social class and race as factors affecting the role of the family in Thematic Apperception Test stories of males. Unpublished doctor's dissertation, Univer. of Pennsylvania, 1950.
34. Mowrer, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.
35. Murphy, G. *Personality*. New York: Harper, 1947.
36. Piaget, J. *The language and thought of the child*. New York: Harcourt, Brace, 1932.
37. Pickering, W. D. A comparison of predominant verbal levels on the Thematic Apperception Test with Rorschach experience-balance. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1950.
38. Piotrowski, Z. A. A Rorschach compendium, revised and enlarged. *Psychiat. Quart.*, 1950, **24**, 543–596.
39. Rapaport, D. On the psychoanalytic theory of thinking. *Int. J. Psychoanal.*, 1950, **31**, 161–170.
40. Rapaport, D. *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.
41. Rorschach, H. *Psychodiagnostics*. New York: Grune & Stratton, 1942.
42. Schachtel, E. G. Projection and its relation to character attitudes and creativity in the kinesthetic responses. *Psychiatry*, 1950, **13**, 69–100.
42. Schumer, Florence. Some behavioral correlates of Rorschach human movement responses. Unpublished doctor's dissertation, Yale Univer., 1949.
44. Shatin, L. Rorschach adjustment and the Thematic Apperception Test. *J. proj. Tech.*, 1953, **17**, 92–101.
45. Siipola, Elsa, & Taylor, Vivian. Reactions to inkblots under free and pressure conditions. *J. Pers.*, 1952, **21**, 22–47.
46. Singer, J. L. Projected familial attitudes as a function of socio-economic status and psychopathology. *J. consult. Psychol.*, 1954, **18**, 99–104.
47. Singer, J. L., & Herman, J. Motor and fantasy correlates of Rorschach human movement responses. *J. consult. Psychol.*, 1954, **18**, 325–331.
48. Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, **16**, 359–364.
49. Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult. Psychol.*, 1954, **18**, 1–9.
50. Singer, J. L., & Sugarman, D. Some Thematic Apperception Test correlates of Rorschach human movement responses. *J. consult. Psychol.*, in press.
51. Singer, J. L., & Wilensky, H. Delaying capacity, fantasy, and planfulness. Paper read at East. Psychol. Ass., Philadelphia, 1955.
52. Thetford, W. N. Fantasy perceptions in the personality development of normal and deviant children. *Amer. J. Orthopsychiat.*, 1952, **22**, 542–550.
53. Vernier, Claire, M., & Kendig, Isabelle V. Analysis of the relationships between various measures of creative productivity in two projective tests. *Amer. Psychologist*, 1951, **8**, 349. (Abstract)
54. Vinacke, W. E. *The psychology of thinking*. New York: McGraw-Hill, 1952.
55. Washburn, M. A system of motor psychology. In C. Murchison (Ed.), *Psychologies of 1930*. Worcester, Mass.: Clark Univer. Press, 1930. Pp. 81–94.
56. Weiss, E. *Principles of psychodynamics*. New York: Grune & Stratton, 1950.
57. Wenger, M. A. An attempt to appraise individual differences in level of muscular tension. *J. exp. Psychol.*, 1950, **29**, 379–390.
58. Werner, H. Motion and motion perception: a study on vicarious functioning. *J. Psychol.*, 1945, **19**, 317–327.
59. Werner, H. Perceptual behavior of brain-injured mentally deficient children. *Genet. Psychol. Monogr.*, 1945, **31**, 51–110.
60. Werner, H. *The comparative psychology of mental development*. Chicago: Follett, 1948.
61. Werner, H., & Wapner, S. Sensory-tonic field theory of perception. *J. Pers.*, 1949, **18**, 88–107.
62. Werner, H., & Wapner, S. Toward a general theory of perception. *Psychol. Rev.*, 1952, **59**, 324–333.
63. Werner, H., & Wapner, S. Organismic theory and perceptual response. Paper read at East. Psychol. Ass., Boston, April, 1953.
64. Wolff, W., & Precker, J. A. Expressive movement and the methods of experimental depth psychology. In H. H. Anderson & G. L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 457–497.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Testing Behavior Hypotheses with the Rorschach: An Exploration in Validation[1]

Arthur E. Eschenbach
Officer Education Research Laboratory, U. S. Air Force

and Edgar F. Borgatta
Russell Sage Foundation

Interest in the Rorschach as reflected in the literature, both in theoretical formulation and report of empirical findings, is extensive. Systematic empirical approaches, however, have been few, and are primarily those involving factor analysis of Rorschach scores (6, 11), or of these scores in conjunction with other scores such as intelligence test performance (10). The scoring categories or diagnostic signs which have been included in the matrix of intercorrelation have not been identical, and where there has been replication of the study, at least reasonable replication of factor structure has been found (10). Factors identified thus far have tentatively been identified with such names as productivity, movement, lack of perceptual control, shading, intelligence, organic pathology, loose conceptual thinking, color awareness and color use, popular response, psychoneurosis, and confused succession. What factors are found is determined in the limit of the approach by the items which are included in the matrix, and the subsections of a population which are included.

Wittenborn (12) has indicated the relevance of a validation study with discrete Rorschach scoring variables, and Wittenborn and Holzberg (13) have presented the complementary argument for using discrete descriptive or behavioral criteria of validation. Essentially, the point is that unless some common variance is found between discrete scoring categories and discrete descriptive or behavioral criteria, there can be little optimism for the diagnostic use of the scoring categories involved. Wittenborn and Holzberg (13) found rather discouraging results in relating Rorschach scoring categories to symptom manifestations in descriptive diagnosis.

Validation of a diagnostic tool should be in terms of the kinds of uses to which it is put. In this study we are interested only in certain aspects of scoring which are defined in terms of the responses of the subject rather than the general impression or evaluation of the interviewer. The use of the Rorschach as a medium of clinical interview is to be distinguished from its use as a diagnostic tool in itself. That is, the Rorschach is a medium of clinical interview when it is used to establish rapport to observe the behavior of the subject or to evoke otherwise pertinent information which might be awkward to elicit in a formal interview. We are not interested in the characteristics which are most closely associated with the sensitivity and intuitive ability of the interviewer, but with the scores which are most arbitrarily defined and reproducible.

---

[1] This research was supported in part by the United States Air Force under contract No. AF 33 (038)–12782 monitored by the Human Resources Research Institute, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication and disposal in whole or in part by or for the United States Government. Personal views or opinions expressed or implied are not to be construed as necessarily carrying the official sanction of the Department of the Air Force. The research was conducted through the Laboratory of Social Relations, Harvard University. A portion of the data utilized was derived from a dissertation submitted by the senior author in partial fulfillment of the requirements for the Ph.D. degree at the University of Florida. We are grateful to the HRRI staff which assisted in various stages of this research.

Procedure

Studies which measure the difference of means of selected Rorschach scores among extreme subsections of the population may prove significant. The information may, however, be of little value for prediction. That is, since two special cells are compared, or one special cell is compared to the remainder of the population, it may be that estimates of the amount of variance actually under control will indicate that the relationship is too weak to be practically useful. In this study we are interested in measuring the relationship of Rorschach scores and behavioral descriptions which exist in a relatively broad population.

Subjects. Our sample consisted of 125 ostensibly normal male Air Force enlisted personnel representing the seven enlisted grades. The subjects were normally distributed on the SRA Test of Primary Mental Abilities, had an average age of 27 years with a range from 20 to 45 years, and were also heterogeneous in regard to education and military occupation. The only apparent nonnormal characteristic of this population was that a majority of the 18 Basic Airmen had recently been incarcerated for breaches in discipline and had lost whatever military grade they had.

Rorschach scoring categories. Since the frequency in some of the commonly used Rorschach scoring categories is extremely small, and the scores are discrete and unitary, these scoring categories cannot properly be utilized in testing relationships among persons. That is, the units of scoring in these categories are too large for the population to be distributed according to ordinary correlational requirements. Split-half reliability of this type of item will almost invariably be negative as a function of the distribution of the scores. For this reason, we have taken scores by the frequency of occurrence without regard to smallest detail specification, e.g., we take C to mean the sum of all cases in which color occurs, such as C, CF, FC, FCV, etc. Our scoring categories in this study are: number of responses (R); $F+/(F+)+(F-)$; total color responses (C); total motion (M); total vista (V); total shading plus texture ($T+Y$); total space (S); total large detail (D); total rare detail (Dd); popular responses (P); organization (Z); *content variety;* animal plus animal detail ($A+Ad$); anatomy (An), total whole (W).

Criteria. We have taken the fifteen scoring categories outlined above and have explored their relationship with five groups of variables selected as valid criteria: (a) scores of performance in interaction; (b) ratings by superiors; (c) primary mental ability scores; (d) sociometric status indices; and (e), background information.

The scores of performance in interaction were taken in a "small group" experiment which is reported elsewhere (4). The scores represent the average number of acts of the person in four sessions which involve getting acquainted, planning role-playing sessions, and free discussion. In each session, the person participated with two different persons. The interaction observed in each session was for a period of 24 minutes, and represented parts of the session in which the subjects were presumably unaware of being observed. The categories of scoring used were those of interaction process analysis (1), which have been used extensively for behavioral description of group and individual performance in discussion and problem-solving groups. Here we include the *initiated behavior* in the twelve basic scoring categories, four subdivisions of the twelve categories, and the total.

Briefly, the *categories of this observation system* may be identified as follows: categories 1, 2, and 3, showing solidarity, tension release, and agreement; categories 4, 5, and 6, giving suggestion, opinion, and orientation; categories 7, 8, 9, asking for orientation, opinion, and suggestion; and categories 10, 11, and 12, showing disagreement, tension, and antagonism. The first three and last three categories are known as the positive and negative social emotional categories, respectively. The central six categories are known as the task categories.

The *ratings by superiors* were made by the highest ranking commissioned officers having direct or effective contact with the airmen. The ratings were in terms of performance variables judged particularly important in the military context: ability to understand the situation and take necessary initiative; ability to handle men and make himself understood by his men; knowledge of the technical aspects of his job; personal conduct and bearing; and general over-all rating.

The scores of *primary abilities* included are those from the SRA Primary Mental Abilities: verbal meaning, space, reasoning, number, word fluency, and a total quotient.

The *affective sociometric indices* are: propensity to be a chooser, propensity to expect to be chosen, choices received (popularity), expectancy of others that one will be a chooser. These were based on a criterion of enjoyment of participation in the session completed. Two other indices on a criterion of amount done to determine the way the session went are also included: chosen as a leader, and chooses self as a leader.

The *background information* includes age, educational level, and the subject's military grade.

Method of Analysis

The method of analysis was governed by experience in the search of empirical hypotheses concerning Rorschach variables.[2] The first question concerned the relationship of R, the number of responses, to the criteria. The second question concerned the relationship of the other Rorschach variables to the criteria. However, since we were primarily interested in the relationships of the other variables not attributable to the number of responses, our matrix was one of partial correlations (Table 1). We were further interested in whether the correlations tend to be ordered properly in regard to behavioral hypotheses which are phrased according to the criteria included. Even if the degree of relationship involved is low, the proper ordering of relationships according to hypotheses would indicate the feasibility of a multiple regression (configurational) approach. Thus we stated forty-four hypotheses. The hypotheses were then tested on the basis of the expected coefficient signs of the relationships of the Rorschach scoring categories with the multiple criteria.

Our definitions of variables for the tests of hypotheses were established before examination of the data.[3] We do not have space for justification of the *behavioral* criteria, which do not have obvious face validity in all cases.

Testing the Hypotheses

In the tests of hypotheses, for economy in presentation, we provide the information in Table 2 in the following order: (*a*) The hypothesis is stated. (*b*) At least one source of the statement is indicated. (*c*) The identification of the variable(s) in Table 1 to which the Rorschach variable is to be compared is stated. (*d*) The *expected* sign of correlation under hypothesis is indicated. (*e*) The *actual* sign of the correlation in Table 1 is given. (*f*) Hypotheses are judged as: rejected (significantly in the wrong direction); in the wrong direction; in the right direction; not rejected (significantly in the right direction). Where two or more possible tests are available, they are given consecutively. For the two-tailed test with .05 level, absolute $r =$ .174 is significant (italicized in Table 1), which is equivalent to the .025 in hypothesis testing.

Summary and Interpretation

Of five hypotheses examined concerning the meaning of the total number of Rorschach responses, one was definitely not rejected and the other four were ordered in the right direction.

Of thirty-nine hypotheses other than those concerning the total number of Rorschach responses, with forty-four possible tests, there were: two nonrejection, twenty-one ordered in the right direction, twenty ordered in the wrong direction, and one rejection.

If hypotheses concerning *intelligence only* are considered, there are (including total R): three nonrejections, four in the right direction, and two in the wrong direction.

While it is interesting to consider statements concerning the need for "configurational" in-

---

[2] Generalizations of relationship of Rorschach scoring categories to personality characteristics have been compiled from the five sources (2, 3, 7, 8, 9) listed in the references. These generalizations have been published separately (5). The generalizations taken for hypothesis test here are those considered most clearly stated in relatively systematic texts.

[3] We are extremely grateful to Robert F. Bales for his help in choosing the interaction behavioral criteria.

Table 1

Intercorrelation of Rorschach Variables with Criterion Variables, Partialled Through *R*

Item	*R*	$\frac{F+}{(F+)+(F-)}$	*C*	*M*	*V*	*T+Y*	*S*	*W*	*D*	*Dd*	*P*	*Z*	Content variety	*A+Ad*	*AN*
Mean of Rx variables	24.4	85.4	2.6	11.1	0.8	2.1	2.4	5.3	15.1	1.6	5.3	23.3	9.2	10.6	1.6
SD of Rx variables	14.4	9.9	2.7	2.6	1.3	2.3	2.8	3.6	10.1	3.7	2.2	17.0	3.9	6.1	1.7
Number of Rorschach responses	1.000	−.085	*.391*	*.443*	*.230*	*.555*	*.547*	*.188*	*.935*	*.723*	*.515*	*.332*	*.805*	*.828*	*.365*
Ratings by superiors:															
a. Ability to understand	.116	−.040	−.003	*.188*	.014	*.219*	.045	.060	−.069	−.036	.127	.170	−.058	−.049	.029
b. Ability to handle men	.061	.014	.071	.128	.049	.094	−.027	.096	−.045	−.041	.170	*.176*	−.100	.020	.151
c. Knowledge of job	.090	−.034	.028	.089	.100	.165	.118	.062	−.094	−.068	.004	.169	−.018	.009	.051
d. Personal bearing	.084	−.068	−.009	.097	.088	.064	−.062	.031	.004	.008	.148	.094	−.070	−.007	*.178*
e. Over-all rating	.111	−.011	.030	*.174*	.063	.120	.020	.071	−.070	−.025	.100	.167	−.065	−.020	.086
S R A primary mental ability:															
f. Verbal meaning	*.174*	.015	*.202*	.166	−.013	.023	.154	.145	−.094	−*.213*	.154	*.275*	.062	−.009	−.027
g. Space	.155	.026	.118	−.053	−.010	.054	.057	.045	−.011	−.095	.008	.088	−.122	.092	−.046
h. Reasoning	.136	.063	.150	.122	.004	.035	*.197*	.172	−*.183*	−.160	.048	*.251*	−.015	.001	−.091
i. Number	*.182*	.041	−.051	.130	.005	.037	−.048	−.084	.040	−.037	−.009	.078	−.152	.054	−.005
j. Word fluency	*.245*	*.193*	.171	.073	.032	.012	.048	*.187*	−.122	−.133	*.200*	*.229*	*.175*	−.063	−.034
k. Total quotient	*.220*	.051	.106	.144	−.006	.023	.068	.142	−.086	−.143	.108	*.214*	−.088	.079	−.087
Sociometric ratings:															
l. Chosen as a leader	.127	−.050	.171	.014	.092	*.221*	.115	*.228*	−.170	−*.187*	−.093	*.223*	.020	−.145	.136
m. Chooses self as a leader	.156	−.054	.128	.047	.071	.036	.069	.131	−.131	−.055	.009	*.178*	*.188*	−*.221*	−.030
n. Choices made (expansiveness)	−.087	.037	−.166	.081	−.118	.004	−*.251*	−.107	.123	*.199*	−.170	−*.183*	−.107	.010	−.117
o. Choices expected	.020	*.175*	−.137	.168	−.042	−.077	−.105	.028	−.022	.083	.044	.016	.103	−.008	−*.227*
p. Choices received (popularity)	.156	−.075	.002	−.061	.042	.070	−.119	.119	−.048	.008	.061	.059	−.126	.152	−.046
q. Expectancy he will choose	.014	−.054	.157	−*.192*	.032	.090	.020	.101	−.065	−.060	.008	−.090	.158	−.026	−.046
r. Age	.014	−.012	*.214*	.004	−.002	.034	.035	*.176*	−.074	−.167	.167	.135	−.021	−.066	.142
s. Education	*.253*	−.062	.044	−.024	.028	.077	.171	−.061	−.044	−.014	−.053	.045	.066	−.038	−.017
t. Present grade	.109	.016	.125	.068	.047	.089	−.040	.107	−.009	−.102	.128	.142	.022	−.018	.086
Rate of initiation in interaction:															
u. 1. Showing solidarity	.032	−.084	−.006	.096	.081	*.258*	−.105	.062	−.017	.027	−.101	.073	−.027	−.032	.055
v. 2. Showing tension release	.050	−.065	−.084	.126	.051	.111	−.009	−.083	.026	.087	−.108	−.025	−*.218*	.065	−.050
w. 3. Showing agreement	.157	.022	.117	.081	.008	*.193*	.087	.162	−.105	−.158	.099	.124	.105	−.095	.054
x. 4. Giving suggestion	*.215*	−.125	.109	.032	.073	*.261*	.096	.114	−.142	−.060	−.105	*.223*	−.035	−.058	−.013
y. 5. Giving opinion	.027	−.059	−.058	−.023	−.014	.042	.035	.038	−.009	−.078	.092	.043	−.059	.017	.151
z. 6. Giving information	.008	−.012	−.004	−.052	.010	−.017	.072	−.027	.021	−.059	.023	.031	.058	−.059	−.006
aa. 7. Asking for information	*.251*	−.126	−.100	−.074	.125	.049	.079	−.053	−.054	.067	−.146	−.003	−.033	.051	.070
bb. 8. Asking for opinion	.142	−.013	.062	.037	.075	.073	.131	.080	−.096	−.119	.014	.165	−.033	.001	.020
cc. 9. Asking for suggestion	.001	−.001	.021	−.004	.014	.019	.012	.009	−.005	−.017	.011	.022	.012	−.010	−.014
dd. 10. Showing disagreement	.124	−.111	.012	−.012	*.193*	*.213*	.087	.005	−.025	−.052	−.060	.084	−.156	.076	−.077
ee. 11. Showing tension	.039	−.062	−.006	−.122	.002	−.096	.063	−.074	.013	.030	.027	−.116	−.036	.072	−.013
ff. 12. Showing antagonism	.069	.017	−.001	*.243*	−.039	−.011	−.112	−.002	−.027	.151	−.018	.054	−*.188*	.041	−.138
gg. Total rate	.138	−.088	−.011	−.017	.052	.121	.114	.026	−.037	−.093	.009	.101	−.027	−.029	.073
hh. Positive affect (1+2+3)	.137	−.034	.036	.134	.045	*.228*	.042	.079	−.063	−.057	−.009	.111	−.048	−.040	.023
ii. Instrumental action (4+5+6)	.047	−.047	−.015	−.039	.008	.046	.073	.014	−.008	−.083	.044	.070	.004	−.036	.061
jj. Instrumental action (7+8+9)	.242	−.116	−.045	−.049	.130	.075	.112	−.010	−.076	.003	−.102	.066	−.029	.038	.053
kk. Negative affect (10+11+12)	.081	−.091	−.003	.096	.057	−.026	.075	−.068	.001	.025	−.089	−.078	−.098	−.095	−.048

Table 2

Sign Test of Hypotheses

Hypothesis	Source	Criteria	Expected signs	Actual signs	Judgment
1. High R is associated with high intelligence	**2**, p. 53; **7**, p. 208	f; g; h; i; j; k	+	+	not rejected
2. High R is associated with compulsiveness, drive for thoroughness	**3**, p. 72	gg, ee	++	++	in right direction
3. Subjects in good humor tend to have a high R	**9**, p. 21	v	+	+	in right direction
4. Sullen and unobliging subjects tend to have low R	**9**, p. 21	dd, ee	++	++	in right direction
5. High R is associated with manic conditions (depression hypothesis is conceived as identical here)	**9**, p. 21	hh, gg	++	++	in right direction
6. High W is associated with high intelligence	**3**, p. 20; **2**, p. 10	f; g; h; i; j; k	+	+	in right direction
7. High W is associated with self-assertiveness	**9**, p. 59	gg	+	+	in right direction
		m	+	+	in right direction
8. High W is associated with compulsiveness	**7**, p. 245	ee, gg	++	−+	in wrong direction
9. High W is associated with manic conditions	**2**, p. 13; **9**, p. 40; **3**, p. 23	hh, gg	++	++	in right direction
10. Low W is associated with anxiety	**3**, p. 23	v, ee	−−	−−	in right direction
11. High W is associated with "overall thinking" (abstract ideas and major issues)	**2**, p. 14; **8**, p. 57	x, y	++	++	in right direction
		h	+	+	in right direction
12. High D is associated with common sense or practicality	**8**, p. 58; **7**, p. 261; **2**, p. 14	a, b	++	−−	in wrong direction
13. High D is associated with depression	**2**, p. 15	hh, gg	−−	−−	in right direction
14. High Dd is associated with low intelligence	**3**, p. 28	f; g; h; i; j; k	−	−	not rejected
15. High Dd is associated with compulsiveness	**2**, p. 14	ee, gg	++	+−	in wrong direction
16. High Dd is associated with nagging and fault finding	**3**, p. 28	ff, x	++	+−	in wrong direction
		dd	+	−	in wrong direction
17. High Dd is associated with low adaptability	**3**, p. 27	a, b	−−	−−	in right direction
18. High S is associated with oppositional behavior	**2**, p. 47; **7**, p. 266; **8**, p. 60	dd, ff	++	+−	in wrong direction
19. Low S is associated with passivity	**2**, p. 47	w, ee	−−	++	in wrong direction
20. Total Z varies directly with intelligence	**2**, p. 12	f; g; h; i; j; k	+	+	not rejected
21. High Z is associated with elation	**2**, p. 13	v	+	−	in wrong direction
22. High $F+$ percentage is associated with attention to task and ability to concentrate	**9**, p. 56	x	+	−	in wrong direction
23. High $F+$ percentage is associated with depression	**2**, p. 19	hh, gg	−−	−−	in right direction
24. Low $F+$ percentage is associated with elation	**2**, p. 19	v	−	−	in right direction
25. High $F+$ percentage is associated with intelligence	**2**, p. 19; **3**, p. 35, 37	f; g; h; i; j; k	+	+	in right direction

Table 2—*Continued*

Hypothesis	Source	Criteria	Expected signs	Actual signs	Judgment
26. High M is associated with high intelligence	9, p. 26; 7, p. 268; 2, p. 23	f; g; h; i; j; k	+	+	in right direction
27. Low M is associated with depression	9, p. 27	hh, gg	++	+−	in wrong direction
28. High M is associated with elation	9, p. 27	v	+	+	in right direction
29. High color is associated with allocentricity and awareness of environment	2, p. 31	bb, cc, x	+++	+++	in right direction
30. High color is associated with excitability	9, p. 98	ee, ff	++	−−	in wrong direction
31. High color is associated with energy and initiative (with which subject attacks problems)	2, p. 31	x	+	+	in right direction
		a	+	−	in wrong direction
		l	+	+	in right direction
32. High color is associated with high affectivity	9, p. 76	hh, kk	++	+−	in wrong direction
33. Low color is associated with depression	2, p. 35; 9, p. 31; 3, p. 45	hh, gg	++	−+	in wrong direction
34. High shading is associated with depression	2, p. 35, 36	hh, gg	−−	++	rejected
35. High shading is associated with passivity	2, p. 35	w, ee	++	+−	in wrong direction
36. High shading is associated with anxiety	3, p. 51	v, ee	++	+−	in wrong direction
37. High vista is associated with the feeling of inferiority	2, p. 33; 9, p. 201	ee	+	+	in right direction
38. High vista is associated with depression	2, p. 34; 9, p. 201	hh, gg	−−	++	in wrong direction
39. Animal form is inversely related with intelligence	2, p. 15	f; g; h; i; j; k	−	+?	in wrong direction
40. High animal response is associated with childhood	8, p. 80	r	−	−	in right direction
41. High anatomy responses are associated with hypochondriasis	2, p. 43; 3, p. 63, 64	ee	+	−	in wrong direction
42. High variety of content is related to high intelligence and high education	7, p. 272; 2, p. 41	f; g; h; i; j; k	+	−?	in wrong direction (but word fluency significant)
43. High variety of content is related to high education	7, p. 272; 2, p. 41, 42	s	+	+	in right direction
44. Low number of popular responses is associated with low intelligence	2, p. 17	f; g; h; i; j; k	+	+	in right direction

terpretation of protocols, these results indicate that if the generalizations implicit or explicit in the Rorschach theory are jointly considered, except in the case of total R, the interpreter will be on quite tenuous ground. The ordering of relationships between the Rorschach variables and the criteria is not dissimilar from chance expectation. Further, not only are most of the relationships of zero order, but the only ones in which this is not the case involve intelligence, which can be approached more efficiently by other instruments.

The implications of findings of this order are twofold: (*a*) While one might question the validity of the criteria, the questioner needs to provide better criteria to demonstrate the relationships exist in fact. Pointing to the significant differences found in carefully preclassified groups is not a demonstration of the validity of the measure. It does not constitute forward selection or prediction. Equally important for consideration is that the preselection criteria may be easier to apply than the test; *extreme* behavior may be recognized by other tools (including even direct questions). (*b*) The matrix of intercorrelations indicates that, although the amount of variance under control in any given instance is small, for limited purposes a forward validation study of the Rorschach may indicate utility of some of the arbitrary scoring categories in use.

Received September 8, 1954.

References

1. Bales, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley, 1950.
2. Beck, S. J. *Rorschach's test.* Vol. II. New York: Grune & Stratton, 1949.
3. Bochner, Ruth, & Halpern, Florence. *The clinical application of the Rorschach test.* (2nd Ed.) New York: Grune & Stratton, 1945.
4. Borgatta, E. F., & Bales, R. F. Task and accumulation of experience as factors in the interaction of small groups. *Sociometry,* 1953, **16**, 239–252.
5. Eschenbach, A. E., & Borgatta, E. F. *A review of some major Rorschach scoring categories and hypothesized personality correlates. USAF Hum. Resour. Res. Inst. Res. Mem.*, 1953, No. 19.
6. Hughes, R. M. Factor analysis of Rorschach diagnostic signs. *J. gen. Psychol.,* 1950, **43**, 85–103.
7. Klopfer, B., & Kelley, D. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
8. Mons, W. *Principles and practice of the Rorschach personality test.* London: Faber & Faber, 1948.
9. Rorschach, H. *Psychodiagnostics.* (4th Ed.) New York: Grune & Stratton, 1949.
10. Williams, H. L., & Lawrence, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.,* 1953, **17**, 261–264.
11. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.,* 1950, **14**, 261–267.
12. Wittenborn, J. R. An evaluation of the use of different scores in prediction. *J. clin. Psychol.,* 1951, **7**, 108–111.
13. Wittenborn, J. R., & Holzberg, J. D. The Rorschach and descriptive diagnosis. *J. consult. Psychol.,* 1951, **15**, 460–463.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Predicting Success on the Member-Employee Rehabilitation Program[1]

Bernard A. Stotsky
VA Hospital, Brockton, Massachusetts

To aid in rehabilitating chronic hospitalized psychotics, the Member-Employee program was instituted in VA-NP hospitals. Selected patients were discharged and placed in full-time salaried positions as employees of the hospital. Employment in a normal work atmosphere was designed to help them make the transition from hospital to community more easily. Selection of member-employees by the Medical Rehabilitation Board was mainly on an intuitive basis.

This study was undertaken to discover objective criteria which could be used to predict success on the program. The Guilford-Zimmerman, Manson, Revised Beta, Purdue Pegboard tests, patient attitude, Palo Alto Hospital Adjustment, Northport rating scales, and information on seven biographical items were collected for the first 18 patients approved by the board. Of all the tests, scales, and items used, only three of the biographical items—*arrests for any except psychiatric reasons, difficulty with alcohol prior to the past two years,* and *present difficulties with alcohol*—differentiated the successful from the unsuccessful member-employees. With the sample increased to 36, differences for the three items were significant at .05 level. The item, *present difficulties with alcohol* classified 35 out of 36 correctly, being significantly related to failure ($p < .001$). Only this item predicted significantly better than the board ($p = .02$). For a second sample of 25 from another VA hospital, the item did not discriminate better than chance though the trend was positive ($p = .11$).

Two main reasons for failure on the program were exacerbation of psychotic symptoms and overindulgence in alcohol. From these findings it appears that alcoholics and incompletely stabilized psychotics are poor prospects for a program of this type.

Brief Report
Received March 9, 1955.

---

[1] An extended report of this study may be obtained without charge from Bernard A. Stotsky, 1233 Warren Ave., Brockton, Massachusetts, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4555 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Relations Among Several Objective Measures of Anxiety Under Different Conditions of Motivation[1]

Anthony Davids
Harvard University

Psychologists have long agreed that the problem of human anxiety is of great theoretical and practical significance, but it seems to have taken the development of a rather simple questionnaire to open the floodgates and release a tidal wave of experimental findings. We will not even attempt to review the host of publications that have appeared on the psychological scene since Taylor recently reported on her scale for measuring manifest anxiety (18, 19). Performance on this scale has been related to a wide variety of experimental measures, including such things as eyelid conditioning (9, 16, 17, 18), verbal learning (11, 15, 20), word associations (3), reactions to stress (4, 22), responses to projective tests (8, 22), and various other complex behaviors (1, 7, 14, 21). Also, a comprehensive review of these studies and a great many others has been prepared (13). So, rather than discuss studies that are probably familiar to anyone acquainted with the recent literature, we will turn immediately to consideration of the present experiment.

In the attempt to discover what unique advantages or disadvantages this new anxiety inventory might have, results were compared on the following instruments: (*a*) Taylor anxiety scale; (*b*) Psycho-somatic Inventory; and (*c*) anxiety self-ratings.

More specifically, there were two main purposes in this study. First, we wished to examine the intercorrelations between results on the different measures within each of two groups of subjects. Second, we wished to see what effect the experimental manipulation of motivation would have on responses to the various objective measures of anxiety.

Method

Subjects. There were two experimental groups, each consisting of 20 male undergraduates selected from an introductory psychology course. With the exception of the experimental manipulation of motivation, described below, the procedure for selecting members of both groups was designed to control as many relevant variables as possible. The groups were matched to the extent where we felt quite certain that any differences revealed on the experimental measures could probably be attributed to the different conditions of motivation under which they were examined.

Treatments. The first group of 20 *S*s completed the questionnaire and self-ratings in the course of their service as paid volunteers participating in an intensive investigation of personality. They believed that the results were anonymous, and were informed that the data would be used solely for purposes of "scientific research" and would not affect them personally in any way. The second group of 20 *S*s was comparable to the first, with the exception that they were informed the results would be used to select one or two members from the group to work as highly paid assistants to staff members conducting research on small group interaction. They were led to believe that we were looking for mature, well-adjusted individuals who would be able to observe and analyze the behavior of others with minimal interference from their own problems and abnormalities. At the time of selection, an interview was conducted with each *S* in order to ensure that only individuals with high motivation to secure the position would be used in the present study.

---

[1] This study, conducted at the Harvard Psychological Clinic as part of a program of research under the direction of Professor Henry A. Murray, was facilitated by the Laboratory of Social Relations. The research project was supported in part by a grant from the Rockefeller Foundation, and by grant M-700 from the National Institute of Mental Health, Public Health Service. The author gratefully acknowledges the assistance of Mr. Leo McNamara, who helped with the scoring and statistical analysis of the data.

Measures. The Taylor scale (19) consists of 50 items selected from the Minnesota Multiphasic Personality Inventory (MMPI) on the basis of clinical judgment that the items are indicative of manifest anxiety. These statements, to which *S* answers either "true" or "false," are embedded in a Biographical Inventory which contains 175 additional buffer items that were also taken from the MMPI. The maximum possible anxiety score is 50 and the minimum is zero.

The Psycho-somatic Inventory (12) consists of 92 questions, of which 46 refer to psychological experiences and 46 refer to physiological experiences. These experiences are supposed to differentiate the "psychoneurotic" from the "normal" individual. The questions are answered on a 4-point scale running from "often" to "never," and the total score based on the entire inventory provides a measure of "neuroticism." High positive scores are indicative of normality and negative scores are indicative of neuroticism.

The anxiety self-rating scale (2) consists of five parts. There are four subratings in response to specifically defined experiences and feelings indicative of anxiety, and one general rating. The four subratings are concerned with (*a*) illness, (*b*) irrational fears, (*c*) insecurity, and (*d*) anxiety symptoms. For each subarea several examples are provided in order to leave little doubt as to the kinds of thoughts and behaviors being assessed. The general rating follows the four subratings, and the instructions read: "Comparing yourself with others of the same sex and age, rate your usual level of general anxiety by placing a cross on the appropriate point on the following scale." Each of the five separate ratings is made on a 6-point scale, running from "much less than most people," to "much more than most people." Summing scores for the five ratings, the maximum anxiety score obtainable is 30 and the minimum is 5.

Table 1

Rank-Order Correlations Among Three Measures of Anxiety Under Two Conditions of Motivation *

Measure	Taylor anxiety scale		Psycho-somatic Inventory	
Anxiety self-ratings	*.60***	.74	*.74*	.84
Taylor anxiety scale			*.76*	.83

* Italicized coefficients are for the 20 "research volunteers" and nonitalicized coefficients are for the 20 "job seekers."

** A coefficient of .52 is significant at the .01 level for a one-tailed test.

Predictions. Our expectation was that there would be a significant positive association between the different measures of anxiety within each group, and that there would be less anxiety evidenced by the ‘job seekers" than by the "helpers of science."

Results and Discussion

The results presented in Table 1 show that within both groups there are significant intercorrelations among the three different measures of anxiety. In both groups, the correlation between the manifest anxiety scale and the Psycho-somatic Inventory is not much lower in magnitude than reliability coefficients reported for the manifest anxiety scale (19). These results indicate that the Taylor scale is not unique. Regardless of the particular instrument employed to assess anxiety, the relative standing of the *S*s is quite similar.

Comparing the mean anxiety scores for the two groups, it is evident from Table 2 that *S*s who are motivated to present a favorable picture of themselves reveal less anxiety in response to each of the three assessment methods. Rather unexpectedly, it is the self-rating scales that show the least difference between the two groups. In response to the Psycho-somatic Inventory, members of the group with nothing to gain by deceiving us reveal themselves as more neurotic. That is, they admit having more experience with disturbing anx-

Table 2

Comparison of Anxiety Scores for Two Groups of 20 *S*s

Measure	"Research volunteers"		"Job seekers"		*F*	*t*
	Mean	Variance	Mean	Variance		
Anxiety self-ratings	15.6	36.1	15.0	36.7	1.00	.29
Psycho-somatic Inventory	137.7	25730.9	202.2	10902.4	2.36*	1.50
Taylor anxiety scale	20.8	64.3	13.8	46.6	1.38	2.96**

* Significant at the .05 level.
** Significant at the .01 level.

iety (lower scores indicate greater disturbance) than do the *S*s who are motivated to appear "well adjusted." The difference between these means does not attain the .05 level of significance, but Bartlett's test reveals that the variance for the group of volunteer research *S*s is significantly greater, which is what we would expect in a group that is not motivated to distort their responses in the direction of social acceptability.

Although the above differences are in accord with our theoretical predictions, it is only the manifest anxiety scale that shows a statistically significant difference between the means for the two groups. This difference of seven points, with the highly motivated *S*s showing less manifest anxiety, is significant beyond the .01 level, indicating that, in the present study, the Taylor scale is especially susceptible to *S*s' conscious or unconscious motivation to present a favorable picture of themselves. This finding receives additional confirmation from the fact that seven *S*s in the group of "job seekers" received manifest anxiety scores that were lower than the score of 10 achieved by the lowest-ranking *S* in the group of "research volunteers."

At this point, let us consider briefly the high degree of consistency among findings derived from different measures of anxiety. This is no isolated finding for several previous investigators have also found significant relations between anxiety measured by the Taylor scale and anxiety measured by other objective instruments. Holtzman, Calvin, and Bitterman (10) correlated Taylor anxiety scores with neuroticism measured by the Winne scale (23), and found a coefficient of .74. When the correlation was corrected for attenuation due to the unreliability of each scale, it increased to .86. These scales represent independent adaptations from the MMPI, and contain 7 overlapping items, but, nevertheless, the magnitude of this relationship is quite high. In another study, Deese, Lazarus, and Keenan (4) found a correlation of .69 between the Taylor scale and Winne scale.

The psychasthenia scale from the MMPI has also been found to relate positively to the Taylor scale. Deese, Lazarus, and Keenan (4) obtained a correlation of .81 between these scales, and a correlation of .40 after overlapping items had been eliminated. Eriksen (5) reported a correlation of .72 between these same scales. And Eriksen and Davids (6) secured a correlation of .92 between psychasthenia and manifest anxiety. This latter finding was reduced only slightly, to .86, when 8 overlapping items were excluded.

These results, coupled with findings in the present study, indicate that performance on the Taylor scale of manifest anxiety is highly related to (*a*) self-ratings on anxiety; (*b*) neuroticism on the Psycho-somatic Inventory; (*c*) neuroticism on the Winne scale; and (*d*) psychasthenia on the MMPI. In many instances, the magnitude of relationship between the independent scales is about as high as reliability coefficients for the Taylor scale. Thus, several seemingly diverse researches seem likely to be investigating very similar, if not identical, phenomena. It may be, as Eriksen and Davids suggest (6), that the Taylor scale is not so much a measure of amount of anxiety as it is a measure of the way people respond to or cope with their anxiety. At any rate, these consistent findings from independent investigations, relating results on the different objective measures to performance in a variety of psychological experiments, suggest that these several lines of research should be brought together and viewed in some common framework.

In conclusion, we would like to point out that the Taylor scale is only one of many available objective techniques for assessing anxiety. And it may well be that other instruments, some of which have been established for years, provide even more valid measures of anxiety than does the Taylor scale. Evaluated according to the criterion of generating research, there is no doubt that this scale has made a notable contribution. Yet the reason why psychologists have seized upon this instrument with such eagerness is not easy to understand. In fact, one could ask, legitimately, "Why the emphasis on the Taylor scale?" An attempt to answer this question is beyond the scope of this paper. Here, we wish merely to stress the point that any one of several objective instruments could probably be employed with equal fruitfulness in the selection and classification of *S*s for investigations of anxiety and behavior.

One thing, at least, seems certain. When objective measures are employed, and responses accepted at face value, the condition under which the *S*s are examined is of considerable importance. When the examinees have something to gain by reporting fewer experiences with disturbing anxiety, these objective measures are quite susceptible to distortion. Results of the present experiment suggest that the Taylor scale may be even more susceptible to deception than are other objective methods for assessing anxiety.

Summary

Two groups of male undergraduates were administered the Taylor anxiety scale, Psycho-somatic Inventory, and anxiety self-rating scale. One group of 20 *S*s took these procedures as volunteers participating in a research project, with anonymity guaranteed. The second group of 20 *S*s was informed that the results of the test battery would be employed to select a few "mature, well-adjusted" individuals to work as paid assistants conducting psychological experiments.

In both groups highly significant intercorrelations were found among the three objective measures of anxiety. When the groups were compared, it was found that on the three measures the group of "research volunteers" achieved scores more indicative of manifest anxiety and maladjustment than did the group of "job seekers." Although these results were all in keeping with our theoretical expectations, it was only on the scale of manifest anxiety that the difference between the two groups attained the .01 level of statistical significance.

It was concluded that (*a*) the Taylor anxiety scale measures essentially the same variable or trait as that measured by several other available instruments, (*b*) the Taylor anxiety scale may be more susceptible to deception than are some of these other instruments, and (*c*) a number of seemingly heterogeneous researches may actually bear upon common phenomena.

Received November 4, 1954.

References

1. Bitterman, M. E., & Kniffin, C. W. Manifest anxiety and "perceptual defense." *J. abnorm. soc. Psychol.*, 1953, **48**, 248–252.
2. Davids, A. The influence of personality on auditory apperception and memory. Unpublished doctor's dissertation, Harvard Univer., 1953.
3. Davids, A., & Eriksen, C. W. The relation of manifest anxiety to association productivity and intellectual attainment. *J. consult. Psychol.*, 1955, **19**, 219–222.
4. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.
5. Eriksen, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–565.
6. Eriksen, C. W., & Davids, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
7. Farber, J. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
8. Goodstein, L. D. Interrelationships among several measures of anxiety and hostility. *J. consult. Psychol.*, 1954, **18**, 35–39.
9. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
10. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's manifest anxiety scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853–854.

11. Lucas, J. D. The interactive effects of anxiety, failure, and intraserial duplication. *Amer J. Psychol.,* 1952, **65**, 54–66.
12. McFarland, R. A., & Seitz, C. P. A psychosomatic inventory. *J. appl. Psychol.,* 1938, **22**, 327–329.
13. Mandler, G. Anxiety and performance: empirical and theoretical aspects of anxiety scales. Unpublished manuscript. Laboratory of Social Relations, Harvard Univer., 1954.
14. Matarazzo, J. D., Ulett, G. S., & Saslow, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.,* in press.
15. Montague, E. K. Role of anxiety in serial rote learning. *J. exp. Psychol.,* 1953, **45**, 91–96.
16. Spence, K. W., & Farber, J. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.,* 1953, **45**, 116–119.
17. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.,* 1951, **42**, 183–188.
18. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.,* 1951, **41**, 91–92.
19. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48**, 285–290.
20. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.,* 1952, **44**, 61–66.
21. Wesley, Elizabeth L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.,* 1953, **48**, 129–134.
22. Westrope, Martha R. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.,* 1953, **48**, 515–524.
23. Winne, J. F. A scale of neuroticism: an adaption of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.,* 1951, **7**, 117–122.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

A Further Investigation of the Relationship Between Manifest Anxiety and Intelligence

Allen D. Calvin, Paul B. Koons, Jr., Joseph L. Bingham, and Howard H. Fink

Michigan State University

Grice (2), who worked with airman basic trainees, recently reported a significant negative relationship between Air Force Clerical Aptitude Index scores (which he states may be regarded as a general intellectual measure) and performance on the Taylor Manifest Anxiety Scale (*A* scale). Kerrick (3), who also worked with Air Force trainees, reported that the *A* scale correlates negatively with a number of Air Force tests of intellectual abilities. Matarazzo *et al.* (4) used college students as *S*s and failed to find a significant relationship between *A*-scale scores and performance on the CVS; however, they did find a significant negative correlation between *A*-scale scores and the ACE. Schulz and Calvin (5) failed in their attempt to replicate Matarazzo *et al.*'s finding of a significant negative relationship between *A*-scale scores and the ACE. Farber and Spence, who have made the most extensive contributions to the *A*-scale literature, state: "The authors have, in fact, been unable, over a period of years, to find any relation between *A*-scale scores of college students and conventional measures of intellectual ability, such as entrance-examination scores and grade-point averages" (1, p. 10).

Because of the widespread use of the *A* scale as a research and clinical instrument, a resolution of these contradictory findings seems called for. The present experiment investigates some of the factors leading to the equivocal results obtained in earlier studies.

Method

Subjects. The *S*s were obtained from two separate sources. The first group, hereafter referred to as Group A, were students in an undergraduate course in psychology at Michigan State University. The group consisted of 36 *S*s, 21 males and 15 females.

The other group, hereafter referred to as Group B, was obtained from a how-to-study course offered for students who were having scholastic difficulty. Two additional *S*s for this group were selected from a class in elementary psychology in which they were experiencing considerable difficulty with the course material. This group had an N of 15, 12 males and 3 females.

Procedure. The Wechsler-Bellevue Intelligence Test (7) was administered in individual sessions to each *S*. The *A* scale was presented with the Biographical Inventory items as suggested by Taylor (6).

Results

The Wechsler scores for Group A ranged from 102 to 141 with a mean of 122.75 and a σ of 9.18. This range, while not unusual for college students, is lacking in low scores; therefore we decided to obtain additional *S*s of lower intelligence, and Group B was obtained (range 92 to 129, mean 108.53, σ 10.41). However in order to obtain *S*s with such low scores it was necessary to select students having trouble with their academic work—this being an almost invariable concomitant of such scores in a university population. This, of course, creates some question as to the propriety of combining the scores of Group B with Group A. We decided to present the data for Group A alone and also to present the findings for the two groups combined. In interpreting the findings for the

Table 1

Pearson *r*'s between *A*-Scale Scores and Wechsler Performance

Item	Group A	Combined groups A and B
Full IQ	−.21	−.31*
Verbal IQ	−.22	−.39**
Performance IQ	−.15	−.18
Information	−.11	−.44**
Comprehension	−.19	−.25
Digit Span	−.37*	−.31*
Arithmetic	−.19	−.31*
Similarities	−.02	−.14
Vocabulary	.08	−.29*
Picture Arrangement	−.02	.13
Picture Completion	−.07	−.19
Block Design	−.36*	−.48**
Object Assembly	−.23	−.29*
Digit Symbol	.02	−.11

* Significant <5% level.
** Significant <1% level.

combined groups (range 92 to 141, mean 118.57, σ 11.56), the reader should, of course, bear in mind the manner in which Group B was selected.

The results for Group A alone and for the combined groups appear in Table 1. The correlations presented are between *A*-scale scores and the derived IQ scores (prorating for vocabulary where necessary), and between *A*-scale scores and raw subtest scores. Examining the findings for Group A first we see that the correlation between total IQ and manifest anxiety is not significant; however, an examination of the subtest correlations reveals significant negative correlations between *A*-scale scores and Digit Span and *A*-scale scores and Block Design.[1]

The results for Groups A and B combined indicate a significant negative relationship between *A*-scale scores and Full Scale IQ, Verbal IQ, Information, Digit Span, Arithmetic, Vocabulary, Block Design, and Object Assembly. Because of the previously mentioned scholastic difficulties of Group B which is confounded with their lower IQ scores, it is impossible to make any unequivocal interpretation of these results.[2] However, these findings indicate the necessity of restricting the conclusions drawn from a particular experiment to the population under study and the specific measure of intelligence used.

Discussion

Although we failed to obtain a significant negative relationship between *A*-scale scores and IQ in our "normal" group of students, which supports the previous findings of Schulz and Calvin (5) and Farber and Spence (1), our finding in the "normal" group of a significant negative correlation between *A*-scale scores and several Wechsler subtests, plus the highly significant results obtained when the students with lower IQ scores (and concomitant scholastic difficulties) were added, indicate the necessity for closely examining the intellectual make-up of the individuals used as *S*s in any experiment utilizing the *A* scale. It seems quite possible, as Kerrick (3) suggests, that the contradictory findings reported by various investigators who used the *A* scale to select high and low anxious groups for learning experiments, conditioning experiments, etc. were due to variations in the intellectual make-up of the respective experimental populations.

Summary

Two groups of college students were given the Wechsler-Bellevue Intelligence Test and the Taylor Anxiety Scale. Group A was composed of 36 students in an undergraduate psychology class. Group B was made up of 15 students with lower IQ scores who were having academic difficulty.

Significant negative correlations between *A*-

---

[1] It is interesting to note that these subtests have been mentioned by numerous clinicians as being good anxiety indicators, and some investigators using different measures of anxiety have also found these subtests able to differentiate anxious *S*s. The experimental findings are far from unequivocal; however, space precludes an adequate discussion of this topic here.

[2] Confounding of another type also exists in the relationship between anxiety and intelligence. As Grice (2) and Kerrick (3) pointed out, it may be that the highly anxious *S* is handicapped on an intelligence test and therefore obtains a lower score than his "true" score.

Our colleague Dr. Charles Hanley suggests another possibility. He feels that the *S*s of lower intelligence may have difficulty understanding the *A*-scale items and may inadvertently check items which raise their *A*-scale scores.

scale scores and several Wechsler subtests were reported for Group A, and when Group B was combined with Group A the *A*-scale scores of the combined groups showed a significant negative correlation with Total IQ, Verbal IQ and a large number of subtests. The implications of these results for contemporary research in this area are discussed.

Received May 9, 1955.
Early Publication.

References

1. Farber, I. E., & Spence, K. W. Main and interactive effects of several variables on reaction time. *U. S. Naval Res. Tech. Rep. No. 3,* Contract N9 onr–93802, State Univer. of Iowa, Mar. 1, 1955.
2. Grice, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.,* 1955, **50**, 71–74.
3. Kerrick, J. S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.,* 1955, **50**, 75–77.
4. Matarazzo, J. D., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.,* 1954, **18**, 201–205.
5. Schulz, R. E., & Calvin, A. D. A failure to replicate the finding of a negative correlation between manifest anxiety and ACE scores. *J. consult. Psychol.,* 1955, **19**, 223–224.
6. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48**, 285–290.
7. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

The Correlation of Intelligence and Neurological Findings on Twenty-Two Patients Autopsied for Brain Damage[1]

Robert S. Morrow
VA Hospital, Bronx, New York

and Joseph C. Mark[2]
VA Hospital, East Orange, New Jersey

This paper is the first part of a more extensive investigation of the effects of various types of structural brain pathology on the psychological test performance of twenty-two adult males. In each of these cases the brain pathology was the main cause of death as determined by autopsy. All of these former patients had received a routine battery of psychological diagnostic tests as well as the usual neurological examination and techniques, i.e., EEG, laboratory studies, etc. The plan of the investigation is to correlate the psychological test results with the neurological and post-mortem findings.

The present paper deals solely with the effect of structural brain pathology on intellectual functioning as measured by the Wechsler-Bellevue Intelligence Scale and specifically it will report on:

1. The relationship between structural brain pathology and total as well as subtest Wechsler-Bellevue scale functioning. That is, the Wechsler-Bellevue results of this obviously brain-damaged group will be analyzed and compared with a control group.
2. Comparing the test scores of those whose locus of damage is in the anterior region of the brain (pre-Rolandic) with those whose lesions are in the posterior regions (post-Rolandic).
3. Comparison between the effects of focal and diffuse brain pathology on the intelligence test.
4. Comparing the effect of dominant cerebral hemisphere lesions with nondominant ones.

Historical Background of the Problem

The literature on the relationship between brain disturbance and psychological functioning is extensive, even when confined to studies of patients rather than animals. Klebanoff (17, 18) has written two excellent surveys. The recent use of intelligence tests such as the Wechsler-Bellevue Intelligence Scale in the studies of brain functioning has yielded valuable information regarding some of the older theories of brain functioning, such as localization theories, and has raised new and important hypotheses. For the most part, however, the studies using the Wechsler-Bellevue test were concerned with the development of test patterns as diagnostic signs as indicated in the writings of Wechsler (27) and Rapaport (21). Allen's (2, 3, 4) studies are typical of this approach.

---

[1] A modified version of this paper was read at the American Psychological Association, New York, September, 1954.

[2] Formerly with the Clinical Psychology Section, Neuropsychiatric Service of the Veterans Administration Hospital, Bronx, New York where this work was done. The authors wish to express their indebtedness to the following: Dr. Louis Berlin, Assistant Professor of Neurology at Cornell University Medical College and former chief of the Bronx VA Hospital Neurology Service, for his assistance on the problems involving neurology, neuroanatomy, and neuropathology; and to the following Bronx VA Hospital staff members: Dr. Jacob Cohen, Assistant Chief Clinical Psychologist, for his help on the statistical evaluations; to Dr. H. L. Flowers, Chief of Neuropsychiatry, and Dr. E. G. Kreuger, Chief of Neurosurgery, for their cooperation and support, and to Dr. Harold G. Wolff, senior consultant in neurology whose searching questions and comments inspired this study.

The major criticism of these studies has been their failure to use adequate control groups to permit the comparison of pathological brain function with the function of persons free from structural brain pathology. Many studies have been reported on the localization of intellectual functions (6, 8, 9, 10, 12, 13, 14, 19, 20, 22, 23, 25, 26) and these have been criticized for failing to use specific anatomical data. This study attempts to take care of both these important objections.

Subjects and Procedures

In this investigation, a brain-damaged group was compared with a control group of psychiatric patients. Each group contained 22 male patients from the Veterans Administration Hospital in the Bronx, New York. The death and post-mortem examination of the brain-damaged patients occurred within the past seven years during which time each had been referred at least once for psychological diagnostic examination from the neurological or neurosurgery services. They were routine referrals handled by different staff and trainee psychologists. Wherever possible each patient received Form II of the Wechsler-Bellevue Intelligence Scale, the Wechsler Memory Scale, the Rorschach, Bender Gestalt, and Figure Drawing tests and in certain instances special tests (i.e., Goldstein-Scheerer, aphasia tests, etc.).

The psychological, neurological, and post-mortem data in this study come from the retrospective examination of the permanent records of the patients. The psychological tests were administered to the brain-damaged patients from one week to twenty-five months prior to death, the median being seven months. The tests were given only when the patient was, according to the judgment of both neurologist and psychologist, in good enough physical and mental condition to be tested.

The same battery of tests was given to the control group from the psychiatric wards who were matched for age, education, and color. The results of this matching as given in Table 1 show that both groups were approximately 40 years old and had about 10 years of formal education. Individual matching for occupation was not considered necessary because most of our patients come from the unskilled occupational groups. The psychiatric patients were either neurotic or mildly schizophrenic but were not at the time of testing receiving any physiological therapy.

Of the group of 22 brain-damaged patients, 18 had direct involvement of the cortex; the remaining 4 consisted of 2 subtentorial (posterior fossa) and 2 third-ventricle tumor cases. Eleven of the 18 cortical cases were considered focal in that the lesion, regardless of size, was circumscribed; these were mostly tumors. The other 7 were considered diffuse in that most or all of the cerebrum was affected, resulting primarily from vascular, toxic, or infectious disease. Of the 18 cortical lesions, 4 were anterior (pre-Rolandic), while 14 were predominantly posterior (post-Rolandic and pre- and post-Rolandic). Also, of these 18, 8 were in the left hemisphere, 2 were in the right hemisphere, and 8 were bilateral. All patients had a history of right-handedness. In making these determinations we were assisted by a neurologist who worked from the autopsy and neurological reports. The autopsy reports contained the microscopic as well as gross brain examination findings of the pathologist.

Results

The Wechsler-Bellevue intelligence test scores for the brain-damaged and psychiatric control groups are presented in Table 1. In addition, the Wechsler-Bellevue scores for 45 normal patients were available and are presented in the last column for further comparison. These patients, tested as part of another study which required normal controls, were recuperating from minor surgery (i.e., appendectomy, tonsillectomy, etc.) and had no history of either brain damage or psychiatric difficulty. They could not be used in this study as normal controls because they had received only the intelligence test and also because they were younger than the brain-damaged group. The data for this group are given for the benefit of those who desire some form of normal reference point other than the standardized norms.

As can be seen from Table 1, the brain-damaged group is significantly lower in Full Scale and Performance IQ, but not significantly lower in Verbal IQ. Most of the subtest scores, with the exception of Vocabulary, Information, Comprehension, and Digits Forward, are significantly lower for the brain-damaged groups. This is particularly true for the performance tasks. The brain-damaged patients fared worst on the Digit-Symbol test, with an average weighted score of 4.23 and four of the subjects making zero weighted scores. On the other hand, the Vocabulary score of the brain-damaged group was the same as the controls, indicating that the vocabulary function in brain damage was apparently unimpaired. On the whole, the low performance test scores together with low Digit Span, Arithmetic, and Similarities corroborated the psychometric test patterns of organic brain pathology as reported in the literature.

Table 1

Wechsler-Bellevue Intelligence Scale Scores of Brain-Damaged and Psychiatric Control Patients

Variable	Mean score: Brain-damaged ($N=22$)	Mean score: Psychiatric ($N=22$)	t*	p	Mean score "normals" ($N=45$)
Age	40.91	39.82	.305	—	30.1
Education	10.00	10.09	.075	—	
Wechsler-Bellevue					
Full Scale IQ	91.95	108.85	3.647	$<.001$	111.9
Verbal IQ	97.67	106.76	1.897	—	110.2
Performance IQ	87.49	109.67	4.804	$<.001$	111.6
Subtest weighted scores					
Information	9.41	10.86	1.571	—	11.5
Comprehension	9.86	11.00	1.356	—	11.2
Digit Span (total weighted)	6.68	9.14	2.081	$<.05$	11.1
Digits Forward (no. correct)	5.73	6.36	1.518	—	7.2
Digits Backward (no. correct)	3.91	5.04	2.211	$<.05$	5.6
Arithmetic	7.09	10.41	2.781	$<.01$	10.5
Similarities	7.64	9.82	2.390	$<.05$	11.1
Vocabulary	11.05	10.89	.173	—	11.2
Picture Arrangement	7.05	10.38	3.112	$<.01$	11.3
Picture Completion	7.32	10.27	3.292	$<.01$	12.2
Block Design	7.00	10.00	2.988	$<.01$	11.1
Object Assembly	7.05	10.36	2.971	$<.01$	10.8
Digit Symbol	4.23	9.14	4.610	$<.001$	10.4

* $df = 35–45$.

Table 2

Wechsler-Bellevue Scores for Cases of Focal, Diffuse, and Subtentorial and Third-Ventricle Lesions

Variable	Focal ($N=11$)	Diffuse ($N=7$)	Subtentorial and 3rd ventricle ($N=4$)	F*	p
Age	37.73	50.71	30.50	5.71	$<.05$
Wechsler-Bellevue					
Full IQ	90.09	87.57	105.25	.56	—
Verbal IQ	95.73	95.71	106.50	.64	—
Performance IQ	85.64	82.28	103.00	1.82	—
Subtest IQ equivalents					
Information	97	98	105		
Comprehension	102	96	109		
Digit Span (Forward & Backward)	86	93	101		
Arithmetic	85	92	103		
Similarities	87	89	106		
Vocabulary	106	104	106		
Picture Arrangement	93	83	100		
Picture Completion	89	92	97		
Block Design	90	85	106		
Object Assembly	90	74	113		
Digit Symbol	70	84	94		

* $df = 2/19$.

Table 2 compares the 11 focal cases with the 7 diffuse ones. In addition, the 4 cases are given in which the lesion did not involve the cortex directly, the 2 third-ventricle and 2 subtentorial tumors. Since there are wide discrepancies in the ages of these groups, the subtest scores are presented as IQ equivalents rather than as weighted scores. The conversion of the subtest scores into an IQ equivalent is achieved on the basis of the standardization tables in Wechsler (27). Because of the paucity of cases, significance tests could not be made.

There is seen to be very little difference between the groups of focal and diffuse lesions, although both groups showed generally lowered test scores resulting from the brain damage. In contrast, the 4 noncortical cases showed, on the whole, very little intelligence test loss when compared with the control group. In both the focal and diffuse groups the Performance IQ is much lower than the Verbal IQ. Among the verbal subtests Vocabulary, Information, and Comprehension did not seem to be adversely affected.

The significance of these findings is that the lowered intelligence resulted from the cortical damage itself and seemed to be independent of the nature of the damage. It mattered little whether the lesion was focal or diffuse when the size of the lesion was somewhat controlled by virtue of the fact that both groups had extensive damage resulting

Table 3

Comparison of Wechsler-Bellevue Scores Between Pre-Rolandic and Post-Rolandic Lesions

Variable	Pre-Rolandic ($N=4$)	Post-Rolandic ($N=14$)	t*	p
Age	39.75	43.64	.53	—
Wechsler-Bellevue				
Full IQ	100.25	85.92	1.62	—
Verbal IQ	105.00	93.07	1.27	—
Performance IQ	94.50	81.42	1.33	—
Subtest IQ equivalents				
Information	103	94		
Comprehension	106	100		
Digit Span (Forward & Backward)	96	81		
Arithmetic	100	87		
Similarities	92	82		
Vocabulary	110	103		
Picture Arrangement	101	89		
Picture Completion	92	86		
Block Design	100	83		
Object Assembly	91	83		
Digit Symbol	79	73		

* *df* = 16.

in death. We could find no typical psychometric test pattern for either the focal or diffuse lesions since both show, on the whole, the typical pattern of gross brain damage.

Table 3 compares lesions anterior to the Rolandic fissure with posterior lesions. Here, too, the subtest scores are given as IQ equivalents because of the age difference. On the whole the patients in the pre-Rolandic or frontal lobe group show only slight intellectual interference, with most of the functions falling within the average range or better, except for the low Digit Symbol score. The intelligence impairment is much greater and much more consistent for those lesions which are either completely or almost completely in the areas posterior to the Rolandic fissure, although the differences are not statistically significant. Here average or better scores were obtained only on Vocabulary, Comprehension, and Information. In both cases, again, the Performance IQ was much lower than the Verbal IQ. The psychometric test pattern for structural brain pathology was much more clearly revealed for the group of post-Rolandic lesions than for pre-Rolandic ones. While few frontal lobe cases were involved in this study, the fact that these four cases show only slight impairment raises questions about some of the special intellectual functions ascribed to the frontal lobes.

Direct comparison between lesions of the left or dominant and right (nondominant) hemisphere is precluded because only two lesions were clearly right-sided and these were located in the frontal lobes. Instead, Table 4 shows the comparison between eight cases of left-hemisphere damage with eight that are bilateral. The two right-hemisphere lesions are also shown but not statistically compared. Despite the lack of statistical significance in the findings, the left-hemisphere lesions showed general loss in Verbal and Performance IQ while the bilateral group showed lowered Performance IQ only. Vocabulary, Information, and Comprehension were relatively unaffected for both groups, but the main differences were in the low Digit Symbol, Digit Span, Arithmetic, and Similarities subtest scores for the dominant hemisphere group, and low Digit Symbol, Picture Arrangement, Block Design, and Object Assembly subtest scores for the bilateral group.

Discussion

The small number of cases in each category precludes broad conclusions or generalizations. We feel justified in making tentative interpretations, however, because of the rarity of the instances where precise localization is available through autopsy. The Wechsler-Bellevue test patterns in this study generally

Table 4

Comparison of Wechsler-Bellevue Scores Between Left-Hemisphere and Bilateral Lesions

Variable	Left hemisphere ($N=8$)	Bilateral ($N=8$)	t*	p	Right hemisphere ($N=2$)
Age	37.12	50.00	2.20	<.06	
Wechsler-Bellevue					
Full IQ	84.12	88.00	.521	—	113.5
Verbal IQ	87.12	99.00	1.591	—	117
Performance IQ	83.25	78.63	.559	—	109
Subtest IQ equivalents					
Information	92	100			115
Comprehension	96	100			120
Digit Span (Forward & Backward)	80	92			108
Arithmetic	73	96			117
Similarities	80	92			101
Vocabulary	102	105			132
Picture Arrangement	95	80			106
Picture Completion	90	91			115
Block Design	89	82			109
Object Assembly	89	75			121
Digit Symbol	71	79			103

* *df* = 14.

Note.—The data for the two right-hemisphere lesions are shown in the last column but since the number is too small, in addition to both lesions being in the frontal lobe, no comparisons are made.

confirm the test patterns of most of the other investigations (21, 27). Some of the slight differences are worthy of comment. For example, Allen's (2, 3, 4) studies of both brain-damaged and brain-diseased patients obtained scores slightly higher than those of this study, but he worked with patients who were presumably healthier since they were seen in an out-patient clinic. Aita and his group (1) compared Wechsler-Bellevue scores of their entire brain-injured population with those classified into mild, moderate, and severe cases. The scores of the severely injured group bear a very close resemblance to ours. It seems, therefore, that the severity or extent of the brain disturbance influences the Wechsler-Bellevue test pattern. The unimpaired Vocabulary score confirms its usefulness as an index of premorbid intellectual capacity as recommended by Babcock (7) and others (21, 27). Other factors which might be involved in the patterning are contingent upon the nature of the disturbance, such as the location of the lesion, etc.

Whether the lesion is focal or diffuse is apparently of very little importance insofar as the effect of severe brain disturbance on the Wechsler-Bellevue test results is concerned. What seems to be of importance is the cortex itself as evidenced by comparisons of cortical lesions with subtentorial and third-ventricle lesions. Our findings seem to be in substantial agreement with those of Anderson (5, 6) who, in comparing what he regards as "focalized" gunshot wounds in the dominant hemisphere with the nondominant ones, reported a higher Wechsler-Bellevue Deterioration Quotient for the dominant-hemisphere cases. He also indicated that cases with dominant-hemisphere damage show a greater loss of verbal abilities while the nondominant-hemisphere cases show a greater loss in performance. McFie and Piercy (19) found that patients with left-sided lesions were inferior on the Weigl Sorting Test to patients with nondominant, right-sided lesions, irrespective of location in the frontal, parietal, or occipital lobes. Milner found that in left temporal lobe lesions causing aphasia the test pattern ". . . is likely to be one of severe deficit on verbal tests and relatively mild deficit on nonverbal ones" (20, p. 55). In our study, six of the eight dominant-hemisphere lesions showed some form of aphasia yet both Performance and Verbal IQ's were low, the Performance mainly because of the Digit Symbol and the Verbal mainly because of the Digit Span, Arithmetic, and Similarities. Weisenberg and McBride (28) showed that right-hemisphere damage resulted in lowered performance tests. Milner indicated that right temporal lesions in particular, as well as right parieto-occipital lesions, interfere with visuomotor functioning, mostly in terms of the perception of spatial patterns. Since there were in our study no right-hemisphere lesions other than the two in the right frontal lobes, the left hemisphere was compared with bilateral lesions. The bilateral involvement was generally similar to the reported findings on right-sided lesions in that the visuomotor tests were lowest, these being Picture Arrangement, Block Design, Object Assembly, and Digit Symbol.

Since the left-hemisphere lesions in this investigation showed both verbal and performance involvement, the question of aphasia merits consideration. Six of the eight dominant-hemisphere cases were reported as having some kind of aphasia, while two of the eight bilateral cases were reported as showing dysarthria but no aphasia. The speech impairment did not seem to interfere directly with the test administration. The fact that the Vocabulary as well as Information and Comprehension scores are either unaffected or only slightly so seems to argue against the language disturbance alone causing the generally lowered intellectual functioning.

Research on cerebral localization has been concerned mainly with frontal lobe function, as in the work of Goldstein (10, 11, 12), Rylander (22, 23), and Halstead (13). Halstead, typically, found that intellectual impairment was greater in frontal lobe lesions than in nonfrontal ones, regardless of whether the lesion was unilateral or bilateral, left or right. Our findings disagree. He found, also, no evidence for mass action by getting no correlation between extent of lesion and degree of impairment. In all, the frontal lobes were considered most essential to biological intelligence with the gradient being from the frontal lobes.

In contrast several studies refute the claims of frontal lobe function. Hebb (14, 15, 16) in his critical review of the work on the frontal lobes, as well as his own evaluation of four cases of frontal lobectomy in which he found no frontal lobe signs, concluded that the evidence on which intellectual defects have been attributed to frontal lobe damage is very poor. Teuber and his co-workers (24, 25, 26) found that problem solving in patients with posterior lesions is at least as difficult as for patients with frontal lobe lesions. The Mt. Sinai group under Bender, consisting of Battersby and his co-workers found that the "intellectual changes noted in patients with tumors of the frontal, temporal, parieto-occipital or even subtentorial areas were about the same in quality, although less frequent in the last instance" (8, p. 705). They stated that "no significant differences between patients with pre- or post-Rolandic lesions could be found" and concluded that "intellectual deficits may be produced by lesions in any portion of the cerebral hemispheres, and not chiefly by frontal lobe involvement, as some authorities are still claiming" (8, p. 712). Aita and colleagues (1) compared the deterioration quotients for anterior and posterior brain-injured patients and found much more "deterioration" for the posterior areas. In our study there was a difference between the frontal lobes and the posterior ones with the greater defects being, in general, in the post-Rolandic areas.

In trying to explain the disagreements among the different studies one should attempt to go beyond the problem of undifferentiated brain disturbance and take into consideration the following factors: (*a*) the use of matched controls; (*b*) the extent of the lesion; (*c*) its location in terms of cortical versus subcortical, dominance versus nondominance, and pre- versus post-Rolandic lesions; (*d*) the nature of the lesion, such as focal versus diffuse, tumors versus vascular disturbances, and disease versus injury; (*e*) the stage of the disease, i.e., chronic versus acute, etc.; as well as (*f*) the kinds of tests used. Other factors and functions which may be of importance in reconciling the discrepancies might be the roles of factors such as aphasia, personality, and visuomotor function. In other words, the total picture of brain damage should be considered.

Summary and Conclusions

The purpose of this paper was to determine the effect of structural brain pathology on intellectual activity as measured by the Wechsler-Bellevue Intelligence Scale, Form II. The results of 22 adult males whose death resulted from various forms of brain pathology were compared with a control group of 22 psychiatric male patients. The autopsy examination reports indicated the nature, extent, and location of the lesions. The following are the main findings:

1. There appears to be a typical Wechsler-Bellevue test pattern for gross structural pathology consisting of significantly lower scores in Digit Symbol, Block Design, Digit Span (especially Digits Backward), Arithmetic, and Similarities. In addition, the Full Scale IQ and the Performance IQ, including all of the Performance subtests, were significantly lower for the brain-damaged population. There were no significant differences involving the Vocabulary, Information, and Comprehension scores. The seemingly unaffected Vocabulary score justifies its use as an indicator of premorbid intelligence.

2. Within the brain-damaged population significant differences were not found between the focal and diffuse lesions, although both groups showed lowered scores. The two posterior fossa and two third-ventricle cases showed practically no intellectual impairment.

3. Under the conditions of this study there is no indication of the so-called frontal lobe signs. On the contrary, patients whose lesions were located anterior to the Rolandic fissure showed far less intellectual loss than those with lesions which were post-Rolandic. The post-Rolandic lesions showed the typical loss and psychometric patterning of gross structural pathology.

4. Comparison between the left- or dominant-hemisphere lesions and bilateral lesions indicated in general that the dominant-hemisphere lesions show significantly lower Verbal IQ (mostly Digit Span, Arithmetic, and Similarities) and Performance IQ (mainly because of Digit Symbol) while the bilateral lesions

show only lowered Performance scores, including Digit Symbol, Block Design, Object Assembly, and Picture Arrangement.

These results are not compatible with the traditional brain theories which emphasized the roles of the frontal lobes and dominant hemisphere in intellectual functioning.

Received November 5, 1954.

References

1. Aita, J. A., Armitage, S. G., Reitan, R. M., & Rabinowitz, A. The use of certain psychological tests in the evaluation of brain injury. *J. gen. Psychol.*, 1947, **37**, 25–44.
2. Allen, R. M. The test performance of the brain injured. *J. clin. Psychol.*, 1947, **3**, 225–230.
3. Allen, R. M. The test performance of the brain diseased. *J. clin. Psychol.*, 1948, **4**, 281–284.
4. Allen, R. M. A comparison of the test performances of the brain-injured and the brain-diseased. *Amer. J. Psychiat.*, 1949–50, **106**, 195–198.
5. Anderson, A. L. The effect of laterality localization of brain damage on the Wechsler-Bellevue indices of deterioration. *J. clin. Psychol.*, 1950, **6**, 191–194.
6. Anderson, A. L. The effect of laterality localization of focal brain lesions on the Wechsler-Bellevue subtests. *J. clin. Psychol.*, 1951, **7**, 149–153.
7. Babcock, Harriet. An experiment in the measure of mental deterioration. *Arch. Psychol. N. Y.*, 1930, **18**, No. 117.
8. Battersby, W. S., Krieger, H. P., Pollack, M., & Bender, M. B. Figure-ground discrimination and the "abstract attitude" in patients with cerebral neoplasms. *Arch. Neurol. Psychiat.*, 1953, **70**, 703–712.
9. Battersby, W. S., Teuber, H. L., & Bender, M. B. Problem-solving behavior in men with frontal or occipital brain injuries. *J. Psychol.*, 1953, **35**, 329–351.
10. Goldstein, K. The significance of the frontal lobes for mental performances. *J. Neurol. Psychopath.*, 1936, **17**, 27–40.
11. Goldstein, K. *After-effects of brain injuries in war.* New York: Grune & Stratton, 1942.
12. Goldstein, K. Mental changes due to frontal lobe damage. *J. Psychol.*, 1944, **17**, 187–208.
13. Halstead, W. C. *Brain and intelligence: A quantitative study of the frontal lobes.* Chicago: Univer. of Chicago Press, 1947.
14. Hebb, D. O. Intelligence in man after large removals of cerebral tissue; report of four left frontal lobe cases. *J. gen. Psychol.*, 1939, **21**, 73–87.
15. Hebb, D. O. Man's frontal lobes: a critical review. *Arch. Neurol. Psychiat.*, 1945, **54**, 10–24.
16. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
17. Klebanoff, S. G. Psychological changes in organic brain lesions and ablations. *Psychol. Bull.*, 1945, **42**, 585–623.
18. Klebanoff, S. G., Singer, J. L., & Wilensky, H. Psychological consequences of brain lesions and ablations. *Psychol. Bull.*, 1954, **51**, 1–41.
19. McFie, J., & Piercy, M. F. The relation of laterality of lesion to performance on Weigl's sorting test. *J. Ment. Sci.*, 1952, **98**, 299–305.
20. Milner, Brenda. Intellectual function of the temporal lobes. *Psychol. Bull.*, 1954, **51**, 42–62.
21. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Vol. I. Chicago: Year Book Publishers, 1945.
22. Rylander, G. Mental changes after excision of cerebral tissue. A clinical study of 16 cases of resections in the parietal, temporal and occipital lobes. *Acta Psychiat. Neurol.*, 1943, Suppl. **25**, 1–81.
23. Rylander, G. Psychological tests and personality analyses before and after frontal lobotomy. *Acta Psychiat. Neurol.*, 1947, Suppl. **47**, 383–398.
24. Teuber, H. L. Some observations on the organization of higher functions after penetrative brain injury in man. In Milbank Memorial Fund (Ed.), *Biology of mental health and disease.* New York: Hoeber-Harper, 1952.
25. Teuber, H. L., Battersby, W. S., & Bender, M. B. Performance of complex visual tasks after cerebral lesions. *J. nerv. ment. Dis.*, 1951, **114**, 413–429.
26. Teuber, H. L., & Weinstein, S. Performance on a formboard-task after penetrating brain injury. *J. Psychol.*, 1954, **38**, 177–190.
27. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
28. Weisenberg, T., & McBride, K. E. *Aphasia: a clinical and psychological study.* New York: Commonwealth Fund, 1935.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

The Awareness of Hostile Feelings in Patients by Their Nurses[1,2]

Sol Charen
Catholic University of America

Subjects were 25 male veteran patients hospitalized for six months and longer with pulmonary tuberculosis. Mean age was 33 years with a range of 20 to 56 years, educational level was at the mean of 10.5 grades with a range of sixth grade to three years of college. Associated with the *S*s were four ward nurses who had daily contact with them during this entire period of hospitalization.

Patients were given the Moldowsky Hostility Test (1) which consists of 25 items from the MMPI dealing with the presence of hostile feelings. The nurses rated the patients through the Factor *H* scale, designed to measure the feelings of hostility in people (2). The correlation between the nurses' scores on the Factor *H* scale and the patients' scores on the Moldowsky was insignificant, − .113.

The experiment was repeated when the patients were transferred to another ward. After three months' stay they again filled out the Moldowsky and the three nurses from this ward rated them with the Factor *H*. The correlation was again insignificant (− .017).

The correlation between the ratings of the first group of nurses and the second group was .579, significant at the .01 level. Mean scores and sigmas for the patients on the test-retest situation were almost identical, with the correlation between tests being .563. The conclusion which the above correlations between nurses' scores and patients' scores suggest is that the nurses seem to have been rating external behavior of the patients although the Factor *H* scale calls for a measure of internal feelings and attitudes. There is obviously better than chance ability on the part of the two groups of nurses to agree in their estimation of a patient's hostility, less however to really know how he feels in this area. Despite close association with the subjects neither group of nurses gives test evidence for empathy.

Brief Report
Received March 29, 1955.

References

1. Dinwiddie, F. W. An application of the principle of response generalization to the prediction of aggressive responses. Unpublished doctor's dissertation, Catholic Univer. of America, 1954.
2. Lorr, M., Rubenstein, E., & Jenkins, R. L. A factor analysis of personality ratings of outpatients in therapy. *J. abnorm. soc. Psychol.*, 1953, **48**, 511–514.

---

[1] From the Veterans Administration Hospital, Baltimore, Maryland. The statements and conclusions reached in this article are those of the author and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2] An extended report of this study may be obtained without charge from Sol Charen, 200 Rhode Island Ave., N.E., Washington 2, D. C., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4554 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Psychological Changes over a Three-Year Period Following Bilateral Prefrontal Lobotomy[1]

Isidor W. Scherer, John F. Winne,
VA Hospital, Northampton, Massachusetts

and Robert W. Baker
Clark University

In a previous publication (1), the authors reported an investigation of the changes in psychological tests of male schizophrenic patients during the first year following prefontal lobotomy. This paper represents an extension of that report to the third year following surgery.

As has been discussed at length elsewhere (1, pp. 1–2), the study was so designed as to take into account several criticisms which may be made of preceding studies of lobotomy: (*a*) failure to control time intervals between successive assessments; (*b*) failure to use control groups as a baseline in terms of which changes in the operated group must be evaluated; (*c*) failure to use a wide variety of psychological measures; and (*d*) failure to integrate test findings with all available clinical data.

Experimental Procedure

Subjects. Fifty white male schizophrenic patients at the Veterans Administration Hospital, Northampton, Massachusetts, have served as subjects during some phase of this three-year study. Of these patients, 28 were lobotomized, the operation consisting of a standard open Lyerly-Poppen approach (1, p. 6) while the remaining 22 constituted the control group.

The subjects for the present study consisted of 20 lobotomized and 16 nonlobotomized patients, fairly well equated for age, education, length of hospitalization, diagnostic classification, and degree of cooperativeness. All patients had received electroconvulsive treatment, which was unsuccessful or only temporarily beneficial, and all had been selected as suitable candidates for lobotomy by the medical staff of the hospital.

Test materials. The experimental materials consisted of a battery of 27 tests from which 113 measures were derived. A description of the tests employed, and the measures derived from them, may be found elsewhere (1, pp. 6–10). Insofar as possible, the test battery was group administered to the operated patients about three weeks before, and two weeks, three months, one year, and three years after lobotomy. The control group was similarly tested at comparable time intervals. Tests with published manuals were scored according to the prescribed procedure; the evaluation of qualitative data was accomplished through group judgment of three clinical psychologists (not including the authors).

Analysis of data. Each of the 113 variables was examined in order to ascertain the statistical significance of the *net* change in the experimental group relative to the control group from testing to testing, using either t tests or net shift in proportion (for dichotomized data). The emphasis, statistically, was on determining the net change in the operated group over and above the change observed in the control group for the particular time period under consideration.

---

[1] From the Veterans Administration, Northampton, Massachusetts. An abstract of this paper was presented to the American Psychological Association, New York, September, 1954. The authors wish to thank Dr. Frank Politzer and Dr. Robert Fortier for their assistance in administration of the test battery and Dr. Cesareo Peña and Dr. Arnold Trehub for critical review of the manuscript.

Problem of the Study

Four major conclusions were derived from the analysis of the data obtained in the one-year study (1). These were: (*a*) There is a tendency toward decreased mental efficiency at one year, associated possibly with organic alteration; (*b*) ego boundaries are strengthened, although a downward trend in the direction of weakened ego boundaries was noted in a comparison of the results at three months and one year after operation; (*c*) sexual awareness and differentiation are increased throughout the entire period, reflecting concern regarding the role which the patient assumes in interpersonal relationships; and (*d*) there is an increase in the rate of action, possibly reflecting a diminution of inhibitory factors. Related to *d*, perhaps, are tendencies toward lack of inhibition in the moral-social field.

The problem in the present study was primarily to ascertain the course of the test measures relevant to these conclusions, with particular emphasis on the period from one to three years following operation. Secondly, the data derived from all of the 113 measures were examined in order to ascertain the changes *within* the experimental group and *within* the control group, as well as the *net change* in the experimental group, for the entire period from before lobotomy to three years after lobotomy.[2] Finally, qualitative material obtained from observation of the patients in several different types of activities and from psychiatrists' progress notes was examined in an effort to determine any additional dimensions of difference between lobotomized and nonlobotomized schizophrenics over the three-year period.

Results

Changes during the period from one to three years following lobotomy. First to be considered are changes in those variables related to the first conclusion from the earlier study.

> There is a predictable pattern of test performance following lobotomy. Performance on tests measuring mental efficiency and organicity (malfunction due to organic factors) drops sharply during the two-week postoperative period. As recovery from the operation takes place, test performance returns to the preoperative base line or rises slightly above it, reaching a peak, in our study, at three months postoperation. There then appears to be a trend toward decline, such that the performance in some tests at one year postoperation falls below the base line, although several tests remain above (1, p. 20).

In the period from one to three years after prefontal lobotomy, a definite pattern suggesting improvement is noted. Of the 29 measures selected as pertinent to this variable, 15 show a shift, at the .20 level of significance or better, in the direction of increased mental efficiency during this time period. There are *no* measures which show a statistically significant net change in the negative direction. This over-all result is significant at the .01 level, when tested with χ^2 as described elsewhere (1, p. 10). The measures, together with the associated p values, are presented in Table 1, where H indicates a change in score signifying better efficiency. Thus, for example, abstraction as measured by the score on the Object Sorting Test is improved from one to three years after operation ($p < .01$); memory is improved, as indicated by increased accuracy in Visual Memory ($p < .05$) and more correct recollections in delayed recall on the Babcock Memory Paragraph ($p < .01$). Thus, the tentative trend toward reduced efficiency which had been observed during the first year is not continued, and, in fact, appears to be reversed.

The second conclusion held that ego boundaries are relatively strengthened one year following lobotomy. However, examination of the changes between three months and one

---

[2] Tables giving, for each measure, the number of cases, means, the standard error of difference, t, and p for the changes within the experimental and control groups and for the net change in the operated group during the period from before operation to three years after operation and for the period from one year after operation to three years after operation have been deposited with the American Documentation Institute. Order Document No. 4556 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Table 1

Direction of Net Change in Measures of Mental Efficiency from One to Three Years Following Lobotomy

Measure	Net change	p
Organic Signs		
Word Association		
Phrases (83)	H	.01
Close (84)	—	—
Personal (87)	—	—
Draw-a-Man		
Shading (69)	—	—
Accuracy (65)	H	.20
Heavy Lines (71)	H	.20
Visual Memory		
Heavy Lines (12)	H	.20
Abstraction		
Categorization (57)	H	.10
Similarities (51)	—	—
Shipley Hartford Rights (55)	—	—
Object Sorting (48)	H	.01
Motor		
Tweezer Dexterity (31)	H	.05
Downey: Total Time (37)	H	.20
Halstead: Right Hand (38)	—	—
Motor Memory		
Halstead		
Both Hands (40)	H	.05
Right-Left (41)	—	—
Positions (44)	H	.10
Memory		
Hard Pairs (9)	—	—
Visual Memory		
Accuracy (10)	H	.05
Memory Par		
Immed. Right (18)	—	—
Delay Right (21)	H	.01
Total Right (24)	—	—
Immed. Distortions (19)	H	.20
Distorted Order (23)	H	.20
Digit Span		
Weighted Score (5)	—	—
Cognitive		
Digit Symbol		
Errors (1)	—	—
Weighted Score (2)	H	.05
Trailmaking		
Errors (59)	—	—
Serial Seven		
Errors (7)	—	—

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates higher mental efficiency, — indicates no change in mental efficiency.

year after operation, although somewhat unclear, suggested that a trend in the direction of weakened ego boundaries had set in. Table 2, which lists the relevant measures, again demonstrates a reversal of trend during the period from one to three years after operation. Of 21 measures studied, 8 are seen to shift, at the .20 level of confidence or better, in the direction of strengthened ego boundaries. One measure, the number of omitted items on Word Association, moves contrary to the general pattern, with an associated p value between .10 and .20. This

Table 2

Direction of Net Change in Measures of Strength of Ego Boundaries from One to Three Years Following Lobotomy

Measure	Net change	p
Memory		
Digit Span		
Forward (3)	—	—
Memory Par		
Immed. Right (18)	—	—
Delay Right (21)	H	.01
Total Right (24)	—	—
Halstead		
Positions (44)	H	.10
Cognitive		
Digit Symbol		
Weighted Score (2)	H	.05
Halstead		
Right-Left (41)	—	—
Right-Both (42)	—	—
Vocabulary (17)	—	—
Body Image		
Draw-a-Man		
Accuracy (65)	H	.20
Contact Measures		
Rorschach		
Rejections (126)	—	—
Multiple Choice Rorschach		
Rejections (138)	—	—
Crude C responses (141)	—	—
Szondi		
p positive (105)	—	—
p negative (106)	—	—
Word Association		
Personal (87)	—	—
Omits (82)	L	.20
Tweezer Dexterity		
Lev of Asp (30)	H	.20
Attitude to God		
Positive (114)	H	.05
Negative (115)	H	.01
Aversion		
Staying at Home (118)	H	.20

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates increased, *L* indicates decreased, — indicates no change in strength of ego boundaries.

Table 3

Direction of Net Change in Measures of Sexual Awareness and Concern from One to Three Years Following Lobotomy

Measure	Net change	p
Bender Gestalt		
Angle Difficulty (77)	H	.05
Draw-a-Man		
Sexual Confusion (66)	—	—
Accuracy (65)	H	.20
Nose Size (64)	—	—
Draw-a-Woman		
Shift in Breast Size (61b)	H	.20
Rorschach		
Reaction Time, Card VI (132)	H	.01
Aversion		
Masturbation (122)	H	.05
Incest (119)	H	.01
Deviant Sexual Behavior (120)	H	.10

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates increased, *L* indicates decreased, — indicates no change in sexual awareness and concern.

over-all result is significant at the .01 level. We see, in Table 2, that measures of memory and cognition improve (p varies between .10 and .01), that more positive ($p < .05$) and fewer negative ($p < .01$) attitudes toward God are expressed, and that level of aspiration is increased ($p > .10 < .20$).

The third conclusion held that there is an increased awareness of sexuality following lobotomy which appears to be maintained for at least three months. Although this increase appeared to be present at one year, the course during the last half of the first year after operation was somewhat equivocal. Table 3, which presents the nine measures pertinent to this variable, indicates that sexual awareness increases during the period from one to three years after lobotomy. The over-all result is presumed to be significant at the .01 level, or better. Seven of the nine measures shift in a positive direction at the .20 level of significance or better. For example, we find increased aversion toward masturbation, incest, and deviant sexual behavior (p varies from .01 to .10), increased reaction time to Rorschach Card VI ($p < .01$), and better delineation of the sex in drawings of human figures ($p > .10 < .20$). Again, there are *no* measures which shift in a direction contrary to the general pattern.

The fourth conclusion maintained that there is an increase in the rate of action in lobotomized patients during the first year following lobotomy—that is, the operated patients tended to act more hurriedly or impulsively than did their controls. Table 4, which lists the direction of net change in the 24 relevant variables, exhibits an interesting phenomenon. Changes in three out of four tests which are primarily motoric (Total Inhibition score of

Table 4

Direction of Net Change in Measures of Rate of Action and Impulsivity from One to Three Years Following Lobotomy

Measure	Net change	p
Motor		
Downey Inhibition		
Total Time (36)	H	.20
$T_3 - T_1$ (37)	H	.10
Finger Dexterity (29)	—	—
Tweezer Dexterity (31)	H	.05
Cognitive		
Digit Symbol		
Weighted Score (2)	H	.05
Serial Seven		
Time (6)	—	—
Word Association		
Time (88)	—	—
Rorschach Reaction Times		
Card III (129)	L	.20
Card IV (130)	L	.05
Card V (131)	L	.20
Card VI (132)	L	.01
Card VII (133)	L	.05
Card VIII (134)	L	.20
Card IX (135)	L	.20
Card X (136)	L	.10
Aversion Items		
Murder (116)	—	—
Theft (117)	L	.20
Incest (119)	L	.01
Deviant Sexuality (120)	L	.20
Withdrawal (121)	—	—
Masturbation (122)	L	.05
Cannibalism (123)	—	—
Expansiveness		
Visual Memory		
Size (11)	L	.01
Draw-a-Woman		
Size (60)	L	.05

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates increased, *L* indicates decreased, — indicates no change in rate of action or impulsivity.

the Downey Will-Temperament Test, the difference between the first and third trials of the Downey, and the O'Connor Tweezer Dexterity Test) suggest that the rate of motoric action continues to increase during the period from one to three years following operation. Perhaps related to this is the increased score in Digit Symbol, a written test which involves speed of motoric action. On the other hand, the data obtained from all Rorschach reaction times, two measures of expansiveness, and four out of seven measures of moral-social attitude suggest that lobotomized patients become less impulsive and less likely to act rapidly in tests involving ideation or imagination.

In summary to this point, we see that lobotomized patients show the following net changes in the period from one to three years after operation: (*a*) a general improvement in tests of mental efficiency; (*b*) a strengthening of ego boundaries; (*c*) a continued increase in awareness of sexuality; (*d*) an increased rate of motoric action; and (*e*) more inhibition or a lesser readiness to act rapidly on tests involving imagination or ideation.

Changes over the three-year period. This paper has been concerned thus far with changes attributable to lobotomy with reference to the period from one to three years following surgery. Next to be considered are the changes over the whole time span employed—i.e., from immediately before lobotomy to three years following lobotomy. Very briefly, tests of significance of the net change in the operated group indicate that only 4 of the 113 measures show impairment which is statistically significant at the .20 level or better: Digit Symbol Errors, Trailmaking Errors, Word Association Omits, and Aversion to Cannibalism. Twenty-three measures show significant net gain, while the remaining 86 measures show no change. Since these results differ from a chance distribution at the .01 level, it may be concluded that the operated patients either held their own or did better on psychological tests over the three-year period than did their controls.

Since the interpretation of many of the 113 measures may be open to question, a group of 40 clear-cut measures of functional efficiency were analyzed in detail.[3] Table 5 presents these measures, the direction of change *within* the experimental and control groups for each of the four testings after operation, and the *net change* in the experimental group over and above the change in the control group. All changes shown are statistically significant at the .20 level or better.[4]

The data presented in Table 5 suggest that the experimental group, except at the second testing, tends to have better efficiency than evidenced prior to lobotomy. For example, three years after lobotomy, gains are shown in 24 tests, no change in 12 tests, and deficit in only 4—increased errors on Digit Symbol ($p < .05$), increased errors on Trailmaking ($p < .10$), and increased confusion in sequence for immediate and delayed recall of the Babcock Memory Paragraph ($p > .10 < .20$). The control group, on the other hand, shows considerable gain when tested for the second time, probably as a result of practice, but by three years has returned essentially to the level attained at the beginning of the experiment.

The final section of the table indicates the relative loss in efficiency of the operated group two weeks after lobotomy, followed by a slow but steady rise over the remainder of the three-year period. We believe that the failure to improve in tests of functioning efficiency—in spite of repeated exposure to the test materials—on the part of the nonoperated group is one aspect of the regressive

---

[3] Tables giving, for each of the 40 measures, the number of cases, means, the standard error of difference, t, and p for the changes within the experimental groups and for the net change in the operated group for the periods (*a*) Pre-2 week post, (*b*) Pre-3 month post, (*c*) Pre-1 year post, and (*d*) Pre-3 year postoperation have been deposited with the American Documentation Institute. Order Document No. 4557 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[4] It will be noted that there is considerable overlap between the measures presented in Table 5 and those presented earlier in Table 1 and Table 2. Table 5 includes all measures adjudged to yield unequivocal interpretation while Table 1 and Table 2 include only those measures which showed significant net change at some period during the three years following lobotomy.

Table 5

Direction of Change from Preoperative Baseline in Forty Tests of Functioning Efficiency over a Three-Year Period

Measure	Experimental group				Control group				Net change			
	2 wk.	3 mo.	1 yr.	3 yr.	2 wk.	3 mo.	1 yr.	3 yr.	2 wk.	3 mo.	1 yr.	3 yr.
Digit Symbol												
Errors (1)	—	—	—	L	H	—	—	L	—	—	—	—
Score (2)	—	H	H	H	H	—	H	—	—	H	—	H
Digit Span												
Forward (3)	—	—	L	—	H	H	—	H	L	L	—	L
Reversed (4)	—	—	—	H	—	H	—	H	—	L	—	—
Weighted Score (5)	—	L	—	H	H	H	—	H	L	L	—	—
Serial Sevens												
Time (6)	—	H	H	H	—	H	—	—	—	—	—	H
Errors (7)	—	L	—	—	—	—	—	—	—	—	—	—
Hard Pairs (9)	—	H	H	H	—	—	H	H	—	—	H	—
Visual Memory Rights (10)	—	—	H	H	—	L	H	L	—	H	—	H
Vocabulary (17)	—	—	H	H	—	L	—	—	—	H	—	—
Memory Paragraph												
Immediate Rights (18)	H	H	—	H	H	H	L	—	H	H	H	—
Delay Rights (21)	—	—	H	H	—	—	H	—	—	—	—	H
Average Rights (24)	—	—	H	H	H	—	H	—	L	—	—	—
Immediate Distortions (19)	L	—	H	H	H	H	H	—	L	—	—	—
Delay Distortions (22)	—	—	H	H	—	—	H	—	—	—	—	H
Average Distortions (25)	—	—	H	H	H	—	H	—	L	—	—	—
Immed. Distorted Order (20)	—	—	L	L	—	—	L	L	—	—	—	—
Delay Distorted Order (23)	—	—	L	L	L	—	—	L	—	—	—	—
Finger Dexterity Time (29)	H	H	H	H	—	H	—	—	H	H	H	—
Tweezer Dexterity Time (31)	—	H	—	—	—	—	—	—	—	—	—	—
Downey Total Time (36)	L	L	—	—	H	—	—	H	L	L	—	—
Halstead												
Right Hand (38)	—	—	H	—	H	H	H	—	L	L	L	—
Left Hand (39)	—	H	H	H	H	H	H	H	—	—	—	—
Both Hands (40)	—	—	H	H	H	—	H	—	L	—	—	—
Shapes (43)	—	—	—	—	—	H	—	—	—	L	—	—
Position (44)	—	—	—	—	—	—	H	—	—	—	—	—
Object Sorting												
Score (48)	—	—	H	—	—	H	H	—	L	L	—	—
Confabulation (49)	—	—	—	—	H	—	—	H	—	—	—	—
Symbolism (50)	—	—	—	—	—	—	—	—	—	—	—	—
Similarities (51)	—	—	—	H	H	—	—	—	L	—	—	—
Series Completion (53)	—	H	H	H	H	—	—	—	L	H	—	H
Shipley Hartford (55)	H	H	H	H	H	H	—	H	—	—	H	H
Categorization (57)	L	—	—	H	—	—	—	—	L	—	—	H
Trailmaking												
Time (58)	—	H	H	—	H	—	H	H	—	—	—	—
Errors (59)	—	—	L	L	H	—	—	—	L	—	L	L
D-a-W Confusion (62)	H	—	H	H	H	—	—	—	—	—	—	—
D-a-M Accuracy (65)	—	H	—	H	—	L	—	—	—	H	—	H
Confusion (66)	H	—	H	H	—	L	L	L	—	H	H	H
Bender-Gest. Elaboration (75)	L	H	H	H	—	L	—	H	L	—	—	—
Word Association Recall (81)	—	—	—	—	—	—	—	L	—	—	—	H
Number of Measures showing												
Gain	5	13	19	24	18	13	12	10	2	10	7	12
No change	31	24	17	12	21	22	23	23	25	23	31	26
Loss	4	3	4	4	1	5	5	7	13	7	2	2

Note.—The number in parentheses refers to the measure as defined in the one-year report (1, pp. 6–10). *H* indicates increased, *L* indicates decreased, — indicates no change in functioning efficiency.

character of schizophrenia. Further study of the apparent differential in this aspect of the regressive process is now underway, utilizing data obtained from the two groups five years after lobotomy.

Additional differences between lobotomized and nonlobotomized schizophrenics. The results which have been presented thus far have been largely quantitative in nature and convey a somewhat favorable connotation for the lobotomized patient. However, qualitative study of the behavior of operated patients and controls during the testing situation, ward activity, and in group and activity therapies suggests that in certain respects the operated group is not doing as well as the nonoperated. Although lobotomized patients appear to be more cooperative, they exhibit a great deal of childishness, lack of motivation, lack of goal-oriented behavior, concreteness of thinking, and irresponsibility—modes of behavior which do not seem to be related directly to the numerous test measures associated with functioning mental efficiency. Moreover, examination of psychiatrists' progress notes does not indicate that there is greater *clinical* improvement in the lobotomized patients than in their controls.

Discussion

The final portion of this report will first present briefly some possible explanations for the findings of increased cooperativeness and efficiency in the lobotomized patients. The apparent discrepancy between test performance and clinical status which was noted in the previous section will then be considered.

First, it appears that the immediate reaction of patients to the operation includes features such as would be found in any individual suffering cerebral trauma. Immediately after lobotomy such typical organic manifestations as increased irritability, concreteness, confusion, temporary release of repression, and a temporary defensive regression were observed. It appears that some of these factors work together to establish a degree of contact with the environment which was absent prior to the operation. The increased contact allows opportunity for social interplay and results in a breakdown of prelobotomy tendencies toward isolation.

A second factor to which we may point for explanation is the commonly observed reduction, following lobotomy, of anxiety associated with pathological preoccupation. This diminution of anxiety seems to remove the need to create new symptoms or to introduce other defensive measures, and so releases psychic energy for employment in the solution of reality-related problems. This could account for increased efficiency on psychological tests.

Third, it is possible that the reduction in the strength and number of morbid symptoms is in part the result of some degree of disorganization or interruption of ingrained thought processes which created the pathological defenses. In addition, the pathological defenses per se become "shaken up" by the organic assault so that they do not operate as effectively. Consequently, the possible deleterious effect of these pathological defenses on functioning efficiency is reduced and, concomitantly, the opportunity is created for new modes of expression of psychic energy.

Despite the increased efficiency noted on psychological tests, qualitative study of the clinical behavior of lobotomized patients does not appear to show the same positive gains. In part, this may be due to lack of clear knowledge of what the tests measure. Although *memory, abstraction, contact,* etc. are defined operationally, there is no certainty about the relationship of the definitions to their corresponding clinical concepts. More importantly, there appears to be no reason why performance on psychological tests should parallel that behavior which is considered in judgments of clinical status, since the latter represent different facets of the total personality and may depend, for measurement, on different variables. Performance on tests seems to depend on intelligence, cooperativeness, and the amount of psychic energy available for use in the solution of the problems which are presented, while judgments of clinical status depend upon a host of variables, many of which are ill defined or in the realm of subjectivity.

Moreover, the apparent lack of gain in clinical status may arise because the lobotomized patient is making more numerous contacts with his environment and is likely to be more demanding on the ward personnel. In a sense, symptoms which are ordinarily encapsulated are more likely to be released and expressed following the operation. Because of this, observers are more apt to notice the operated patient's psychotic ideation and behavior than those of his less demanding and less active control and, consequently, adverse clinical judgments may be made.

Summary

This study deals with the changes in psychological tests in a group of schizophrenic patients during the three years following lobotomy, with particular emphasis on the period from one to three years following operation. The subjects consisted of 20 male white lobotomized schizophrenics and 16 nonlobotomized controls, all fairly well equated for age, education, length of hospitalization, diagnostic classification, and degree of cooperativeness.

All lobotomized subjects were group tested about three weeks before, and two weeks, three months, one year, and three years following lobotomy. Controls were similarly tested at comparable time intervals. The test battery consisted of 27 psychological tests from which 113 relatively independent measures were derived. The data obtained from each of the 113 measures were examined to ascertain the significance of the *net change* in the experimental group in the period from one to three years following lobotomy. Secondly, statistical tests were made of the differences *within* the experimental group, *within* the control group, and of the *net change* in the experimental group for the period from prior to operation to three years after operation.

The results are as follows:

1. Lobotomized patients show the following net changes in the period from one to three years after operation: (*a*) a general improvement in tests of mental efficiency; (*b*) a strengthening of ego boundaries; (*c*) a continued increase in awareness of sexuality; (*d*) an increased rate of motoric action; and (*e*) more inhibition or a lesser readiness to act rapidly on tests involving imagination or ideation.

2. Lobotomized patients generally did better on psychological tests over the three-year period than did their controls.

3. Lobotomized patients, except at the testing two weeks after lobotomy, tend to have better efficiency than that evidenced prior to lobotomy.

4. Control patients show considerable gain on the second testing, but at the end of the three-year period have returned essentially to the level attained at the start of the experiment. It is suggested that this lack of improvement in spite of practice on the tests is one aspect of the regressive character of schizophrenia.

5. Qualitative analysis of the lobotomized patients' behavior and examination of psychiatrists' progress notes suggest that in some respects the operated group is not doing as well clinically as the nonoperated group.

6. It is suggested that the increased efficiency on psychological tests evidenced by the operated group is due to (*a*) increased contact, (*b*) lessened anxiety which releases psychic energy for use in the solution of reality-related problems, and (*c*) disorganization of the pathological defenses as a result of the organic assault.

7. It is suggested, also, that the discrepancy, in the operated group, between psychological efficiency and clinical behavior may be due to (*a*) ill-defined test and clinical measures and (*b*) release of formerly encapsulated symptomatology which may encourage adverse judgments by ward observers.

Received November 10, 1954.

Reference

1. Scherer, I. W., Winne, J. F., Clancy, D. D., & Baker, R. W. Psychological changes during the first year following prefrontal lobotomy. *Psychol. Monogr.*, 1953, **67**, No. 7 (Whole No. 357).

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Diagnosis of Organicity by Means of Spiral Aftereffect

A. Cooper Price
VA Center, Biloxi, Mississippi

and H. L. Deabler
VA Hospital, Gulfport, Mississippi

One of the persistent problems facing a clinical psychologist working in a hospital setting is that of having a reliable and valid test for determining organic brain damage. Recent reports on the more frequently used tests such as the Rorschach, Wechsler-Bellevue, and others tend to cast some doubt on their ability for consistent differentiation of organic from nonorganic cases (9).

In 1949, Freeman and Josey (1) published a study in which ability to perceive a certain spiral aftereffect by patients with memory impairment was investigated. They rotated an Archimedes spiral for a period of time after which rotation was stopped. Normal subjects experienced a visual negative afterimage or effect of either expansion or contraction of the spiral immediately after the rotation was stopped. This apparent movement aftereffect has been referred to elsewhere as gamma movement (3). Freeman and Josey reported that most of their subjects (*S*s) with clinically judged memory impairment did not report or were unable to perceive this distinctive aftereffect.

The results of their study led the present authors to believe that this phenomenon might be utilized for development of a technique for detection of organic brain involvement. It was hypothesized that nonorganic patients would be able to perceive the distinctive aftereffect while organic patients, especially those with cortical involvement, would be unable to perceive it or if they did perceive it, their perception of it would be incomplete. It was decided that research should be conducted to verify this hypothesis and, if results were favorable, to develop this technique as an aid for diagnosis of organicity.

Method

Apparatus. An electric motor and stand commonly used for color mixing experiments was used to produce the necessary rotation. To the shaft, an opaque disk of white plastic 6 inches in diameter was attached. An Archimedes spiral of 920° or 2½ circuits about the center was constructed of dark colored plastic and was placed against the white disk background. The spiral was reversed for the effect of opposite rotation. The disk was turned at approximately 100 revolutions per minute. It was later found that a spiral could be rotated on a phonograph turntable of 78 rpm with identical results.

Procedure. Testing took about five minutes per *S* and was conducted only in the daytime when illumination was good. The *S* was seated approximately 8 feet from the apparatus. Standard instructions were given immediately preceding the test as follows: "This is a special eye test (so stated in order to get attention and cooperation from psychotic *S*s). Look at the center (pointing) and don't take your eye away until I tell you to." Then the rotation was begun. Spiral A, giving normal *S*s a negative aftereffect of expansion, was presented first in half of the cases for 30 seconds. This was ample time for normal and other *S*s. While the disk was rotating *S* was asked "What does the line appear to be doing?" This was done to make sure *S* was attending and could report his experience. After 30 seconds, the disk was stopped and *S* was then asked immediately, "Now what does the line appear to be doing?" The patient's answer was recorded. The routine was repeated with the spiral reversed (Spiral B), giving normal subjects a negative aftereffect of contraction. This was followed by a repetition of presentation of A, then B for a total of four trials of 30 seconds each. Each normal report, that is, seeing the negative aftereffect correctly, was scored 1. Each abnormal report, that is, failure to perceive the negative aftereffect, was scored 0. Reports of aftereffect in which no apparent change of dimension occurred but in which apparent motion forward or backward was reported, were given a score of ½. Also, a doubtful or equivocal report was scored ½. The *S*s scored a total of 4, 3, 2, 1, or 0 depending on the degree of their perception of the aftereffect. In the case of normals the negative aftereffect was so immediate and vivid that it was scarcely necessary to conduct the usual inquiry for scoring. There was little ambiguity as to the presence or absence of the aftereffect with most *S*s. In the interests of conservatism, to give *S* the benefit of a doubtful response, fractional or half scores were raised to whole scores in final computation.

Strict adherence to the procedure for the testing of *S*s was exercised. It was believed that variations from the standard directions might have considerable bearing upon the validity of the results obtained. Precaution was taken not to suggest or structure in any way the nature of the perception being investigated.

Table 1

Aftereffect Score

Subject group	Perception				
	None	Partial			Complete
	0 Score	1	2	3	4 Score
Normal ($N=40$)	0 0%	1 2.5%	0 0%	2 5%	37 92.5%
Nonorganic psychiatric ($N=40$)	0 0%	0 0%	1 2.5%	1 2.5%	38 95%
Organic ($N=120$)	72 60%	13 10%	24 20%	9 8%	2 2%

In order to demonstrate the perceptual aftereffect to be cortical phenomenon and not retinal (2), several normal *S*s were tested with one eye covered. At the end of the 30-second exposure period, the rotation was stopped. At that time the *S*s were instructed to cover the eye which had been used to view the rotating spiral and to use the other to view the now motionless disk. It was found that the aftereffect occurred just the same.

Subjects. The VA Hospital at Gulfport, Mississippi and the VA Center at Biloxi, Mississippi, supplied the *S*s for this study. Three groups were used:

1. Forty normal adult males employed at the institutions at the time of testing, with no known history of organic involvement.
2. Forty nonorganic psychiatric cases. The diagnoses of these were based on hospital staff action. The group was composed of schizophrenics, manic-depressives, and neurotics. Patients recently or currently undergoing electroconvulsive treatment were not used in the study, since it is known that this form of treatment may produce temporary organic effects (4).
3. One-hundred and twenty organic cases with known cortical involvement. Diagnoses of these cases were based on hospital staff action. This group included cases of CNS syphilis, cerebral vascular accidents, chronic brain syndrome associated with alcoholism, cerebral arteriosclerosis, post-traumatic encephalopathy, and advanced stages of Parkinson's disease.

Results

Scores of *S*s in the three groups and percentage computations are given in Table 1.

The chi-square technique was used to determine the possibility of a statistically significant difference between the organic and the nonorganic cases. Chi square for these data was found to be significant at the .001 level of confidence.

Discussion

Functional psychiatric cases were presumed to be without organic brain damage. The test data confirm this hypothesis, cases in this group being able to perceive the aftereffect to a marked degree. The normal and the functional psychiatric *S*s showed no significant difference in their performance.

Standlee recently published an article (7) on the spiral aftereffect as an indication of memory impairment with conclusions differing from those of Freeman and Josey (1) and with implications from our study. He administered the spiral test to 16 normal and 25 psychotic *S*s and found that all were able to perceive the aftereffect. He also gave these same *S*s the Wechsler Memory Scale and found that their Memory Quotient scores ranged from below 48 to above 143, lower and upper limits of the scale, with low scores indicating poor memory ability and high scores indicating superior memory ability. He reasoned from this wide variation in memory ability scores and from the aftereffect report of each of his *S*s that "Freeman and Josey's 'strong probability of correlation' between the visual phenomenon and memory ability is not confirmed" (7, p. 317). Standlee has, however, made a crucial error in his reading and interpretation of Freeman and Josey's study, for they specifically stated that they were attempting to deal with memory *impairment*, not with variations in general memory ability. A distinction exists and must be made between these two conditions. A given individual may have low memory ability and may not have suffered from any type of brain damage or memory impairment. One of our nonorganic psychiatric *S*s, for instance, was a primary mental defective, and he was able to perceive the aftereffect completely.

In the second part of his study, Standlee administered the same tests to his psychotic *S*s approximately eight hours after their receiving one electric shock. On the spiral retest, he found two reporting incomplete aftereffect perceptions with 23 reporting both expanding and contracting perceptions. In his retesting for evidence of memory impairment, he used the same form of the Wechsler Memory Scale as previously used and found the average MQ score for the group actually higher than before shock. On the basis of these data, he concluded that electric shock did not impair memory ability or the ability to experience the spiral aftereffect.

This second part of Standlee's study is also open to criticism. First, his results do show that electric shock had some effect on perception of the aftereffect since two of his postshock *S*s reported incomplete perceptions, and it is open to question as to how many of the other twenty-three drew upon or utilized their previous experience with the test in giving their reports on retesting. Practice effect may have been operative here as well as elsewhere. Second, use of a single shock likely had minimal effects. Had more been used, results would likely have been more marked as has been shown by other studies (4, 8). Third, as he himself admits, the use of the same form of the Wechsler Memory Scale for evidence of impairment is faulty because of practice effects, thus leaving him with no adequate criterion of memory impairment. His study would have fared better had he used in his postshock testing Form II which is available and which was used effectively by Stone (8).

While we did not set out to verify the relationship between memory impairment and the spiral aftereffect, our findings and conclusions, for reasons stated above, are more in agreement with those present in the study of Freeman and Josey than those of Standlee. We are in agreement with Standlee's study insofar as we both found psychotic *S*s—ours specifically diagnosed as functional or nonorganic—as being able to perceive the aftereffect. His study, however, did not deal adequately with the impairment factor which may be of organic or of functional origin. If impairment exists and if it is of organic origin, we predict difficulty or inability to perceive the spiral aftereffect.

In summary, our data for the organic cases contrasted significantly with the nonorganic in the direction of expectancy. When failure to perceive the aftereffect is found, it is almost certain that cortical involvement exists.

The explanation of our significant results in psychophysical terms has not yet been worked out or established. Some leads may be found in Saucer's article on processes of motion perception (5) and in Shapiro's investigation on perception of apparent motion (6). Saucer is inclined to believe that motion perception is derived from a higher-level integrating or synthesizing process in the central nervous system, this integrating or synthesizing process appearing to be adversely affected or nonexistent in individuals with organic brain involvement. Shapiro believes that the perception of apparent motion is related to an "irradiation of excitatory effects" (6, p. 58) and that there is "an exaggeration of inhibitory effects in brain damaged patients" (6, p. 67) which weakens or destroys motion perception. While adequate explanation of the phenomenon remains to be hypothesized and verified, the results are such as to justify present clinical use of this technique in diagnosis of organicity.

Conclusion

Organic cases with cortical involvement can be differentiated from nonorganic with high degree of certainty by means of the spiral aftereffect technique.

Received November 19, 1954.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–796.
2. Gibson, J. J. The visual perception of objective motion and subjective movement. *Psychol. Rev.*, 1954, **61**, 304–315.
3. Graham, C. H. Visual perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 868–920.
4. Kalinowsky, L. B. Organic psychotic syndromes occurring during electric convulsive therapy. *Arch. Neurol. Psychiat.*, 1945, **53**, 269–273.
5. Saucer, R. T. Processes of motion perception. *Science*, 1954, **120**, 806.
6. Shapiro, M. B. A preliminary investigation of the effects of continuous stimulation on the perception of "apparent motion." *Brit. J. Psychol.*, 1954, **45**, 58–67.
7. Standlee, L. S. The Archimedes negative aftereffect as an indication of memory impairment. *J. consult. Psychol.*, 1953, **17**, 317.
8. Stone, C. P. Losses and gains in cognitive functions as related to electro-convulsive shocks. *J. abnorm. soc. Psychol.*, 1947, **42**, 205–215.
9. Yates, A. J. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, **51**, 359–380.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

The Efficacy of Diagnostic Pattern Analysis with the Wechsler-Bellevue[1,2]

Jacob Cohen
VA Hospital, Montrose, New York and City College, New York

Since the publication of the revised manual for the Wechsler-Bellevue scales (5), and more particularly, since the publication of the Menninger group (4), clinicians have attempted to draw inferences about personality and clinical diagnosis from the intersubtest differences of Wechsler-Bellevue records. This practice, variously referred to as pattern analysis, scatter analysis, and profile analysis, has many enthusiastic proponents in the clinical field. It has developed, in a remarkably short time, an enormous research literature with results which have been characterized by Magaret as "negative, conflicting, or questionable" (3, p. 370).

The evidence for the validity of pattern analysis comes mainly from researches in which the scores of criterion groups of patients are compared with those of controls, subtest by subtest. Significant differences discovered by t test are harvested and are then offered as differentially diagnostic signs.

The evidence against the validity of pattern analysis comes from several kinds of investigations: (*a*) subtest-by-subtest comparisons like those cited above which have failed to produce significant differences; (*b*) applications of previously adduced patterns or signs on new samples with failure to substantiate them; (*c*) reworking of data previously offered as indicative of differentiating patterns to demonstrate that, in fact, no such patterns exist.

The approaches of the researches described to the problem under question are indirect. The presence of significant differences between groups on some subtests is no guarantee either that the matching of the groups has been complete enough to remove the effects of extraneous variables, or that these differences are large enough to lead to fruitful applications on individual cases. The absence of significant mean differences does not preclude the possibility that two obverse patterns, either of which is identifiable by a clinician as pertaining to a single diagnostic group, may not cancel each other out and lead to differences which fail of significance. Wittenborn's (6) demonstration with Rapaport's data that pattern correlations within diagnostic groups are no larger than those between diagnostic groups—in the present author's opinion the clearest and most definitive attack on the problem—still leaves a logical loophole in that the clinician may differently interpret the same pattern when the level of the constituent scores varies. In all probability, this loophole does not loom very large from a practical viewpoint. Still, Wittenborn's approach to the problem is an indirect one.

The Problem

When directly stated, the question which is basically at issue is, "Can experienced clinicians, using only the Wechsler-Bellevue subtest weighted scores and the verbal, performance, and full-scale IQ's, make nosological diagnoses of individuals which are significantly related to neuropsychiatric criterion diagnoses?" If they can, a second question of degree of agreement arises and is of practical importance.

---

[1] From the Clinical Psychology Section, Neuropsychiatric Service, Veterans Administration Hospital, Bronx, New York.

[2] The author wishes to acknowledge the cooperation of Dr. H. L. Flowers, Chief Psychiatrist, and Dr. Robert S. Morrow, Chief Clinical Psychologist, of Bronx VA Hospital. He also appreciates the efforts of the seven clinical psychologists who served as subjects.

Materials

In the course of another investigation (1, 2), Wechsler-Bellevue data (eleven subtest weighted scores and the three IQ's) for 300 successive admissions meeting research criteria were gathered. These data were made up of groups of 100 cases each of neuropsychiatrically diagnosed psychoneurotics, schizophrenics, and brain-damaged patients. These were all white male veterans, aged 20 to 40, with unambiguous diagnoses which classified a patient into one and only one of the above three nosological groupings. The three groups were equated for mean and standard deviation with regard to age, years of education, and full-scale IQ. The frequency of subcategories in each group and the means and standard deviations of full-scale IQ, age, and years of education are given in the report of the earlier investigation (2).

Method

Subjects. The subjects proper were seven clinical psychologists with from four to fourteen years of diagnostic experience with the Wechsler-Bellevue. Two of the subjects had published research on the Wechsler-Bellevue.

Procedure. The 300 patterns were randomly ordered and submitted to each psychologist individually with the request to classify each pattern as either psychoneurotic, schizophrenic, or brain-damaged. The subjects were informed that there were 100 of each but that they need not maintain this equality in their diagnoses.

For each psychologist contingency tables were set up as follows: (*a*) A 3 × 3 table was set up to determine the over-all nonchance accuracy of diagnosis between the psychologist's diagnoses and the neuropsychiatric criterion diagnoses. This type of evaluation is illustrated for one of the psychologists (A) in Table 1.[3] (*b*) To determine the nonchance accuracy of psychologist's diagnoses with regard to each diagnostic group separately, 2 × 2 tables were set up which effected a correlation between psychologist's diagnoses and the criterion diagnoses for each group with the other two combined. Table 2 illustrates this type of correlation for the psychoneurotic group utilizing the diagnoses of Psychologist A. (*c*) To determine the degree of relationship between pattern and criterion diagnoses, where chi square was significant, the corrected contingency coefficient was computed for the 3 × 3 table and phi coefficients for the 2 × 2 tables.

Table 1

Degree of Agreement Between Psychologist A's Wechsler-Bellevue Pattern Diagnoses and Neuropsychiatric Criterion Diagnoses for the Categories of Psychoneurotic, Schizophrenic, and Brain-Damaged

	Neuropsychiatric criterion			
A's diagnoses	Psychoneurotic	Schizophrenic	Brain-damaged	Total
Psychoneurotic	56	39	28	123
Schizophrenic	24	33	29	86
Brain-damaged	20	28	43	91
Total	100	100	100	300

Note.—$\chi^2 = 20.11$, $df = 4$, $p < .001$, $C_c = .31$.

Table 2

Degree of Agreement Between Psychologist A's Wechsler-Bellevue Pattern Diagnoses and Neuropsychiatric Criterion Diagnoses in Discriminating Psychoneurotics

	Neuropsychiatric criterion		
A's diagnoses	Psychoneurotic	Other	Total
Psychoneurotic	56	67	123
Other	44	133	177
Total	100	200	300

Note.—$\chi^2 = 13.95$, $df = 1$, $p < .001$, $\phi = .22$.

Results

Table 3 gives the results of the chi-square tests for the seven psychologists in terms of chi-square values. The first column of the chi-square values pertains to the 3 × 3 tests for each psychologist. The second, third, and fourth columns refer to 2 × 2 tests for psychoneurotics versus other, schizophrenics versus other, and brain-damaged patients versus other, respectively.

---

[3] Psychologist A was selected for use as an example of how the chi-square tests were performed. His was the most successful performance, and the reader would be seriously misled if he thought him typical. On the contrary, his attempt at prediction illustrates the inadequacy from a practical viewpoint of even the best performance in this group of experienced clinicians.

Table 3

Chi-Square Values of Significance of Contingency Between Psychologist's Wechsler-Bellevue Pattern Diagnoses and Neuropsychiatric Criterion Diagnoses

		2×2			
Psychologist	3×3	Psycho-neurotics vs. other	Schizo-phrenics vs. other	Brain-damaged vs. other	Number "correct"
A	20.11**	13.95**	1.38	11.38**	132
B	21.18**†	2.11	1.98	14.52**	115
C	8.44	.36	2.33	5.13*	117
D	2.41	.52	.57	1.92	111
E	6.74	.19	.01	.19	107
F	1.20	.06	1.05	.07	94
G	4.54	.75	.03	1.47	102

* $.05 > p > .02$.
** $p < .001$.
† Reflects in part some *in*correctness.

Table 3 immediately indicates that Psychologists D, E, F, and G failed to agree with the neuropsychiatric criterion to an extent greater than chance expectation. This holds for both the over-all (3 × 3) contingency as well as for the treatment of the three groups separately (2 × 2).

Psychologist A, with 132 "correct" diagnoses (chance expectancy is 100 "correct") showed extra-chance agreement with the neuropsychiatric criterion in the over-all analysis ($p < .001$). In addition, he discriminated psychoneurotics from others and brain-damaged from others significantly ($p < .001$).

Lest the reader be overly impressed by the magnitude of the chi-square value and its high order of significance, stress should be placed on its exact meaning. It indicates that one can be quite confident that the hypothesis of no association is false; it does *not* mean that the *degree* of association is strong. The contingency coefficient for the 3 × 3 test for Psychologist A, corrected for the number of categories, was .31. For the psychoneurotic vs. other and brain-damaged vs. other associations, A's phi coefficients were, respectively, .22 and .19. This obviously leaves much to be desired in the way of accuracy.

Psychologist B's results may be misleading. Although the over-all contingency is very significant ($p < .001$), part of this large chi-square value is attributable to the fact that the number of "correct" diagnoses for schizophrenics is *below* chance expectancy. This discrepancy adds to the value of the chi square but, of course, detracts from the over-all correctness of the diagnoses. This is evident from the fact that he has a total of only 115 "correct" diagnoses, which is not significantly better than chance.

This is a good point at which to emphasize the fact that chi square measures the degree of departure of a set of empirical frequencies from some hypothetical set, and does not distinguish the direction of departure. When the hypothetical set of frequencies reflects the hypothesis of a chance association (the null hypothesis), departures in *both* directions, i.e., both less than chance and in excess of chance, equally increase the chi-square value. It is therefore necessary, when using this method, to carefully scrutinize the contingency table cell by cell, before concluding that a significant chi square means better than chance prediction.

Psychologist B does identify brain-damaged patients with above chance success ($p < .001$), but with a small degree of relationship, i.e., phi = .22.

Psychologist C, with 117 "correct" diagnoses, does not attain better than chance agreement with the criterion in the over-all comparison. However, he, too, identifies brain-damaged patients with better than chance success ($p < .05$), but with a phi coefficient of only .13.

Discussion

If the theoretical question is posed, "Is there any detectable association between a neuropsychiatric diagnostic criterion and diagnoses made from patterns?" the answer must be in the affirmative. If the results of the less ambiguous 2 × 2 tests are considered it can be seen that four of these are significant, three of them at the .001 level. This result is unlikely to be accounted for by chance alone. It is also unlikely to be a chance phenomenon that three of these four significant relationships occurred with regard to the brain-damaged group, or that of the seven psychologists, three could significantly discriminate brain-damaged patients from others, while only one succeeded with the psychoneurotics and none with the schizophrenics. The conclusion that some experienced psychologists can make diagnoses from Wechsler-Bellevue patterns significantly better than chance is inescapable. It is also likely that the success they enjoy comes largely from the ability to discriminate significantly brain-damaged patients from others. This is the positive side of the picture.

If a more practical question were posed, namely, "How useful are Wechsler-Bellevue patterns for neuropsychiatric diagnosis?" the answer, on the basis of the present results, must be, "Not very." Four of the seven psychologists did not fare better than chance; indeed, one of them fell below mean chance expectancy. The most successful subject, A, made only 132 "correct" diagnoses out of the 300, and thus was correct 43 per cent of the time, not quite 10 per cent more than the 33⅓ per cent expected by chance. If due allowance is made for the 100 "correct" chance diagnoses, then this psychologist made 32 nonchance "correct" diagnoses for the remaining 200 cases and can be credited with correctly diagnosing 16 per cent of the cases on a nonchance basis.

The small degree of predictive accuracy is also apparent from the magnitude of measures of relationship employed. The best of these, a corrected contingency coefficient of .31 for A's 3 × 3 contingency, presents problems in interpretation, but is in any event quite low. The phi coefficient, which is a product-moment coefficient and hence interpretable approximately as would be a Pearson *r*, gave values ranging from .13 to .22 in the four instances of better-than-chance association.

That Wechsler-Bellevue pattern analysis cannot serve as a sufficiently valid independent diagnostic technique is so obvious that it is not worthy of comment. The data suggest, however, that even as a supplementary diagnostic aid it is of questionable value. For many psychologists (four out of the seven well-qualified subjects of the present investigation), impressions derived from the pattern should be given no weight in the diagnostic decision. For some, it should play so small a role as to be of little consequence.

Summary

Seven experienced clinical psychologists attempted to diagnose the Wechsler-Bellevue patterns of 300 neuropsychiatrically diagnosed veteran male patients as psychoneurotic, schizophrenic, or brain-damaged. Using neuropsychiatric diagnoses as criteria, it was found that only one clinical psychologist correctly diagnosed a significant number of these 300 patients, and only two others had extra-chance success in the diagnosis of a single diagnostic group (in both cases, the brain-damaged group). The actual rate of success, however, ranged in correlational terms from .13 to .22 and was deemed far too small to be of use clinically.

It is concluded that although there is some nonchance relationship between Wechsler-Bellevue pattern and neuropsychiatric diagnosis, it is detectable for only some clinicians and then only to so small a degree as to be of little practical utility.

Received December 16, 1954.

References

1. Cohen, J. A factor-analytically based rationale for the Wechsler-Bellevue. *J. consult. Psychol.*, 1952, **16**, 272–277.
2. Cohen, J. Factors underlying Wechsler-Bellevue performance of three neuropsychiatric groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 359–365.
3. Magaret, Ann. Intelligence testing and clinical practice. In L. A. Pennington & I. A. Berg (Eds.), *An introduction to clinical psychology*. New York: Ronald Press, 1948. Pp. 353–379.
4. Rapaport, D. *Diagnostic psychological testing*. Vol. I. Chicago: Year Book Publishers, 1945.
5. Wechsler, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkiams, 1944.
6. Wittenborn, J. R. An evaluation of the use of Bellevue-Wechsler subtest scores as an aid in psychiatric diagnosis. *J. consult. Psychol.*, 1949, **13**, 433–439.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Specific Ability in IQ Change[1]

Charles T. Baker, Lester W. Sontag, and Virginia L. Nelson[2]
Fels Research Institute, Antioch College

Many investigators have found marked individual differences in the form of the mental growth curve when each individual's intelligence quotients or mental ages are plotted over a period of years.[3] Cycles of slow or rapid mental growth covering periods of long duration appear to be characteristic of large numbers of children from preschool years to adolescence. The consistency and magnitude of the trends of change in mental growth rate, repeatedly found in longitudinal studies (3, 4, 5, 6, 8, 10), indicate that these changes are not due to statistical artifacts. To date, the factors which contribute to accelerative and decelerative changes in mental growth rate have not been clearly defined.

In the study presented here we have investigated whether groups showing accelerated trends, decelerated trends, or little change in mental growth rate differ in their ability to pass the different types of Stanford-Binet items. Insofar as individual items of the Stanford-Binet require somewhat different abilities for their solutions, differences among groups in item difficulty provide information about group differences in areas of specific ability. The items of the Stanford-Binet have often been used in this diagnostic way. McNemar (9) suggested that groups of certain items could provide rough estimates of ability in nonverbal, memory, and vocabulary types of items.

Information regarding the group differences in ability to pass different types of items is also useful in determining the importance of a test-construction artifact in IQ changes over long periods of time. Since the Binet test includes more so-called nonverbal or performance type items at preschool levels and contains more verbal items at later ages, children with greater verbal ability could presumably score lower at earlier ages and higher at later ages as a function of the changing composition of the test itself, while the converse could be true of children whose best abilities were in the nonverbal area. This hypothesis has sometimes been given to account for the inconstancy of the IQ and the low predictive value of early tests (1, p. 294; 7, p. 643).

Procedure

The data for 140 children from the Fels Research Institute longitudinal study were used in the present analysis. Fifty-eight of the Fels subjects (*Ss*) admitted relatively early in the program received the 1916 Revised Stanford-Binet test at birthdays from age 3 until age 6; while 82 *Ss* admitted to the program in later years received Form L or Form M of the 1937 Revision at birthdays and six months after birthdays from age 2½ to age 6. From age 6 to 12 all 140 children received either Form L or Form M at birthdays. Form L and Form M were usually presented alternately. With few exceptions the tests were administered within a few days of scheduled birthdays or half-birthdays, and were administered by one examiner throughout the testing program. The mean IQ of the Fels children at any age is considerably above average on a variety of intelligence tests, the mean Stanford-Binet IQ for this sample being near 120 at all ages with standard deviations varying from 14 to 20 IQ points (10).

---

[1] Presented in part at the Annual Meeting of the American Psychological Association, New York, September 1954.

[2] The authors gratefully acknowledge the many helpful suggestions of Dr. John I. Lacey and the statistical assistance of Mrs. Josephine D. Lubing.

[3] When one refers to measures of gain or loss in IQ trends over long periods of time this is equivalent to referring to measures of accelerative or decelerative changes in mental growth rate. Whether one chooses to graph IQ's or MA's is purely a matter of whether one prefers constancy to be represented by a straight line or by the diagonal line $y = x$. Accelerative and decelerative changes are depicted relative to the line which is chosen to represent constancy.

In order to divide the total group into experimental groups of those with accelerative changes, decelerative changes and constant mental growth rate trends, the following procedure was used. A smoothed longitudinal IQ curve was prepared for each *S* by the method of moving averages. Prior to age 6 smoothing was done by obtaining the mean of blocks of three consecutive raw scores when the interval between tests was six months, and by blocks of two when the interval between tests was one year. After age 6, blocks of three raw scores were averaged, the interval between tests being one year. The purpose of the smoothing was to eliminate erratic variations and thereby effectively portray general trends of IQ change. The different forms of the test were not taken into account in averaging the raw scores.

Quantification of the general trend in IQ change for each *S* was made in the following way. For the *S*s who had been tested at yearly intervals prior to age 6 a "base IQ" was taken to be the mean IQ of raw score IQ's obtained at ages 3 and 4. For those who had been tested at six-month intervals the "base IQ" consisted of the mean IQ of tests given at ages 2½, 3, and 3½. For each *S* the amount of change in IQ from the "base IQ" to the highest smoothed IQ (in the case of those with accelerated trends) or to the lowest smoothed IQ (for those with decelerated trends) was determined. This yielded a gain or loss amount relative to the "base IQ." The *S*s were then ranked from the individual gaining the most to the individual losing the most IQ points. The ranked sample of 140 *S*s was then divided into experimental groups.

Group A consisted of the highest ranking 25 per cent of the sample, that is, those with greatest accelerative changes in mental growth rate. Group B, the 25 per cent of the sample who ranked in the middle of the distribution of ranks, was the group which showed least change; and group C was selected as the 25 per cent of the group who ranked as losing in IQ. It may be noted that only 75 per cent of the total sample was used. Approximately 13 per cent of the cases which ranked between group A and group B were eliminated from statistical treatment as were approximately 12 per cent of the cases ranking between groups B and C. These cases were discarded after graphic inspection of the individual IQ curves showed that changes in mental growth rate in these two parts of the ranked distribution were more erratic than many of the cases falling into the other groups.

The total sample was also subdivided in another dimension, being split into two matched groups (group I and II) of 70 each for purposes of cross-validation in this study and in future studies of these same *S*s. Matching variables were date of birth, sex, level of IQ at age 3, and amount and direction of maximum IQ change. Matching was accomplished in such a way that groups A, B, and C were subdivided approximately equally so that statistical analysis of differences among groups A_I, B_I, and C_I could readily be cross-validated with groups A_{II}, B_{II}, and C_{II}.

A system devised by Baldwin (2) was used for estimating difficulty in first passing items of the Stanford-Binet, Form L. The difficulty level for an individual in first passing an item was defined as the mental age halfway between the mental age obtained by the individual at the time of his first success on the item and his mental age at the time of his previous failure on the same item. From this information about the difficulty of any item for each individual, it was possible to estimate the mean difficulty level of each item for the experimental groups.

Statistical analysis consisted of *F* tests of the differences in mean difficulty level among the experimental groups. This item analysis was done for 60 items of Form L of the Revised Stanford-Binet, and consisted of all the items from the test level of year IV through the year-XII series of items.

Results and Discussion

Seven of the 60 *F* tests of differences among groups A_I, B_I, and C_I were found to be significant below the .10 level in the first analysis. The seven items of the Stanford-Binet, Form L, which resulted in these significant differences, were as follows: at year IV, items 2 and 6; at year IV–6, items 1, 3, and 6; at year V, item 2; and at year VI, item 5. None of the differences on these same items were significant upon replication with groups A_{II}, B_{II}, and C_{II}. In the second analysis 30 F tests were run on those items which had the greatest differences in mean difficulty level between groups A_I, B_I, and C_I. There were no trends toward significance between the first and second analysis. Since item difficulty was made relative to mental age, and therefore controlled for the effects of not only differential mental growth but also IQ level, it is concluded that there are no differences among the experimental groups in ability to pass different types of items. In view of the negative results on Form L, no attempt was made to analyze the data from Form M.

For the experimental purpose of this study it appeared to make no difference whether different forms or one form of the Stanford-Binet was used to obtain the smoothed trend line of IQ change. Smoothed IQ curves based on Form-L IQ's were prepared for those *S*s who had received the 1937 Revision with Form L and Form M alternately presented. Using the procedure of ranking the individuals according to amount of change from the "base IQ" resulted in the *S*s being placed in the same experimental groups as before. Smoothed curves based on Form-M IQ's only yielded similar results also. A comparison of the distributions of gains and losses was also made between those who had received the 1916 Stanford-Binet during the preschool years and those who had received the 1937 Revision throughout the testing period. There were no significant differences between these two groups in the mean amount of points gained or lost from the "base IQ" to the point of greatest divergence.

The maximum number of *S*s possible in each experimental group in the first analysis was 18. The use of the Baldwin technique in estimating item difficulty level resulted in loss of data in the analysis of some items below the VII-year level of the test. Group A_I was combined with group A_{II}, B_I with B_{II}, and C_I with C_{II} for a nonreplicated design in testing two items when the number in one experimental group fell below 10. This analysis resulted in no significant differences in item difficulty.

Conclusions

Within the limitations of the sample and methods of this study it is concluded that there is no evidence of differences in ability to pass any type of Stanford-Binet item between those who accelerate in mental growth rate and those who decelerate. Accelerative or decelerative change in mental growth rate, at least in higher levels of IQ, is apparently associated with similar changes in growth rate in more specific areas of ability. There is no evidence that change in the rate of mental growth reflects a spurious gain or loss in IQ due to an artifact of test construction, wherein the ratio of verbal to nonverbal abilities changes with chronological age. These conclusions must, of course, be limited to the IQ level dealt with in this study.

It would appear that the factors which are associated with individual differences in the mental growth curve are more likely those factors which influence learning in any situation. Sontag and his associates (10) have presented data on the mental growth careers of the Fels subjects, and have proposed various aspects of motivation as determining factors in IQ change. Further research in this area promises to contribute knowledge of those factors which determine the form of the individual's mental growth curve.

Summary

The major purpose of this study was to determine whether *S*s with accelerative changes in mental growth rate differed from those with decelerative changes in any specific areas of ability, as measured by the difficulty in first passing items of the Revised Stanford-Binet Scale, Form L. Matched experimental groups were obtained from 140 Fels *S*s of normal to very superior intelligence, who had been tested at regular intervals from age 2½ to age 12. Cross-validation procedure was used in analyzing 60 items. There were no trends toward statistically significant differences between the groups in this analysis.

It is concluded that at least in the higher levels of intelligence, children with accelera-

tive changes in mental growth rate do not differ in any specific areas of intellectual ability from those with decelerative changes in mental growth rate. There was no evidence in this study to support a hypothesis that changes in Stanford-Binet IQ from the preschool years to later years are due to the changing ratio of verbal to nonverbal items in progressive age levels of the test.

Received December 16, 1954.

References

1. Anastasi, Anne, & Foley, J. P. *Differential psychology*. New York: Macmillan, 1949.
2. Baldwin, A. L. The relative difficulty of Stanford-Binet items and their relation to IQ. *J. Per.*, 1948, **16**, 417–430.
3. Bayley, Nancy. Mental growth in young children. *Yearb. nat. Soc. Stud. Educ.*, 1940, **39** (II), 11–47.
4. Bradway, Katherine. IQ constancy on the Revised Stanford-Binet from the preschool to the junior high school level. *J. genet. Psychol.*, 1944, **65**, 197–217.
5. Ebert, E., & Simmons, K. The Brush Foundation study of child growth and development. *Monogr. Soc. Res. Child Develpm.*, 1943, **8**, No. 2 (Whole No. 35).
6. Hilden, A. H. A longitudinal study of intellectual development. *J. Psychol.*, 1949, **28**, 187–214.
7. Jones, H. E. The environment and mental development. In Leonard Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1954. Pp. 631–696.
8. Katz, E. The constancy of the Stanford-Binet IQ from three to five years. *J. Psychol.*, 1941, **12**, 159–181.
9. McNemar, Q. *The revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin, 1942.
10. Sontag, L. W., Baker, C. T., & Nelson, Virginia L. Personality as a determinant of performance. Paper read at Amer. Orthopsychiat. Ass., New York, March, 1954.

Journal of Consulting Psychology
Vol. 19, No. 4, 1955

Rigidity on the Rorschach and Response to Intermittent Photic Stimulation[1]

Laverne C. Johnson and John A. Stern
Department of Neuropsychiatry, Washington University School of Medicine and Research Laboratories, Malcolm Bliss Hospital

Most theories which attempt to account directly or indirectly for certain major phenomena of human behavior subscribe to a "unified" concept of personality. Such theories assume a high degree of constancy in the manner in which people perceive and respond to various situations. However, the experimental evidence on which such theories are based is meager and of recent origin (9, 11, 18). The experiment to be reported in this paper is, we hope, a contribution to the validation of one of the basic tenets of such theories of personality, namely, that a person will respond in a relatively constant manner to different perceptual situations. In addition, this study offers new opportunity to explore the mechanisms or processes which underlie responses to the Rorschach method as well as to show that it can be effectively used as an instrument for the study of perceptual processes (2, 10, 14).

The perceptual hypothesis to be investigated in this experiment is the following: A person who responds to one set of stimuli in a "rigid" or "flexible" manner will respond to other stimuli in a similar manner. While recognizing the importance of understanding the motivation of behavior, we are, at the present time, more concerned with determining that there are characteristic perceptual "styles" or "attitudes" which remain constant from one situation to another. Demonstration of the latter is a necessary first step toward the study of motivation.

In this study we are specifically concerned with the constancy of perceptual "flexibility" and "rigidity" from one ambiguous perceptual task to another. While well aware of the confusion surrounding the term "rigidity," we believe that it comes closer to describing one extreme of the process we have in mind than other less equivocal terms. By perceptual rigidity we mean that the person responds in a relatively restrictive or selective manner to environmental cues, as compared to the flexible person who responds in a less restrictive and selective manner to these same cues. Clinically, this type of person is often characterized as constricted and one who finds it necessary to deal with his environment in an unusually guarded and stereotyped manner.

The two perceptual tools utilized in this investigation are the Rorschach method and a photic stimulation experiment. In evaluating data from the Rorschach, Fisher's measure of "personality rigidity" was used (6). This measure is based on ". . . a variety of Rorschach signs which clinically have been found to characterize persons who are habitually restricted and who find it necessary to deal with the environment with an unusual amount of guarded caution" (6, p. 9). We feel that this definition, at least in part, describes what we mean by perceptual rigidity. Fisher's scale also has been found to be related to other measures of perceptual rigidity (1).

---

[1] This study was supported in part by a research grant from the National Institutes of Mental Health of the National Institutes of Health, United States Public Health Service.

These data were gathered under Air Force Contract Number 33(038)–13884 with the Department of Clinical Psychology, United States Air Force School of Aviation Medicine.

We wish to express our appreciation to Dr. George A. Ulett for making the data available and for his helpful suggestions in preparation of this paper.

The second ambiguous perceptual situation was intermittent photic stimulation at various frequencies. In this experiment the subject was requested to report what he saw after stimulation at various frequencies ranging from 3 to 30 flashes per second. These subjective sensations were the data used in this study.

In both tasks, the subject is asked to state what he experienced and what the stimulus material reminded him of or looked like. In both tasks the stimuli are ambiguous and permit a variety of responses.

In studies utilizing intermittent photic stimulation, Ulett and associates (15) reported various subjective sensations experienced by subjects. The content of these sensations ranged from pure description of colors and geometric designs to organized illusions and three-dimensional perceptions. Some subjects reported feelings of being projected into space, spinning, dizziness, anxiety, nausea, or other marked dysphoric experiences. Ulett found that it was possible to discriminate fairly well between nonanxiety-prone and anxiety-prone individuals according to the kind of sensation experienced during photic stimulation. More feelings of dysphoria were reported by anxiety-prone individuals.

As part of a larger study reported elsewhere (16), each subject was administered a Rorschach and his subjective sensations were recorded during intermittent photic stimulation. It was thus possible to obtain measures of perceptual rigidity for each task. Substantiation of the hypothesis that rigidity on the Rorschach is related to response to photic stimulation will add further support to the growing belief of constancy of perceptual "styles," as well as elicit new evidence that Rorschach responses reveal adaptive processes which are not different in kind from those appearing in other situations.

Subjects and Procedure

Subjects were 117 male students drawn from the entering class of a local Protestant seminary. The age range of the group was from 18 to 26 with a median of 20. For each subject (*a*) a battery of psychological tests, including the Rorschach, was administered, (*b*) an interview by two psychiatrists was given, and (*c*) an electroencephalogram (EEG) was recorded. Part of the EEG procedure included a photic stimulation survey of 3–30 flashes per second. There was a 40-second exposure to each of 21 different frequencies. Photic stimulation was produced by a pendulum-type episcotister interrupting a tungsten-filament light source which projected a 30-inch circle of light of 200 foot-candles intensity on an opal glass screen placed directly before the subject and filling his visual field. After each 40-second exposure, the subject was asked what he had seen or experienced. His responses were recorded as nearly verbatim as possible. These verbalizations were the data used in this study.

The psychological battery was the first procedure in the research; thus the Rorschach always preceded exposure to photic stimulation by at least a day. All the Rorschachs were administered and scored by the same person, an intern psychologist receiving training at Washington University School of Medicine after having completed all course requirements for the Ph.D. degree. The Klopfer scoring system (12) was used except for scoring of form level and originals where Hertz' (8) scores were used.

The measures of perceptual rigidity. As mentioned above, Fisher's rigidity scale was used as the measure of perceptual rigidity manifested in the Rorschach. The criteria for scoring the subjective sensations were a modification of those used by Klein and Schlesinger (10) in their study of perceptual attitudes toward instability. These criteria were developed and the subjective sensations were scored independently of the Rorschach measure of rigidity. The criteria used in this study were:

1. Criteria for perceptually rigid group.
   a. Little verbalization; verbalization and elaboration does not increase as flicker frequency increases; verbalization may tend to decrease.
   b. Responses are detailed descriptions of stimulus, e.g., geometric patterns or designs, lights, blocks, spots, circles, dots, squares, etc.
   c. Movement seen predominantly descriptive and related to stimulus, e.g., whirling, shimmering, coming and going, up and down, fluctuates, fades in and out.

  - *d.* Uses same percepts, either color, form, or motion, and is not able to shift or elaborate on responses even when flicker frequency increases.
- 2. Criteria for the flexible group.
  - *a.* Much verbalization and elaboration which tends to increase as flicker frequency increases.
  - *b.* Responses not stimulus bound nor descriptive in nature, allows self to project forms and objects into stimulus.
  - *c.* Movement seen also not stimulus bound nor predominantly descriptive, allows self freedom in expression and can take attitude of "distance" from stimulus.
  - *d.* Content not stereotyped, uses more than one type of percept, allows self freedom to elaborate and change responses as flicker frequency increases.

Subjective sensations of each subject were judged as perceptually rigid or perceptually flexible by both authors and also were rated as to how well they met the criteria for that group. All ratings were made on a four-point scale of confidence extending from positive judgment that the subject falls into the given category to the feeling that the judgment was extremely tenuous. There was no fixed weighting procedure and it was rare that all the criteria for one group were met. The rating was in terms of the general over-all impression, and each judge was free to determine the importance of any one criterion. Disagreement between the two judges as to the group in which a subject should be placed eliminated the subject from the study. A difference of more than two points as to how well a subject met the criteria for the group in which he was placed also eliminated the subject. Fifteen subjects were eliminated because of difference as to group placement, and twelve more were eliminated because of difference of more than two points in the rating of certainty that the subject met the criteria. This reduced the N from 117 to 90. The over-all agreement between the two judges was 77 per cent, agreement for group placement alone was 87 per cent.

Fifty-six of our subjects were classified as rigid and thirty-four as flexible on the basis of their subjective sensation to photic stimulation.

Results

Table 1

Comparisons of Response to Photic Stimulation and Rorschach Rigidity Score

	Rorschach rigidity score				
Photic stimulation	N	Mean	SD	CR	p
Perceptually rigid	56	33.1	11.7	4.79	.0001
Perceptually flexible	34	20.8	12.8		

The results of the comparison of the rigidity score on the Rorschach and response to photic stimulation are presented in Table 1. These results clearly indicate that, in this sample, rigidity on the Rorschach is related to our measure of perceptual rigidity in response to photic stimulation. Those subjects who were placed in the perceptually rigid group in terms of their response to photic stimulation had significantly higher Rorschach rigidity scores. A biserial correlation was computed to obtain an estimate of the relationship between the two perceptual measures. This correlation was .59. If a Fisher Rorschach Rigidity score of 25 is used as a cutting point, eight of the flexible subjects (24 per cent) would be incorrectly placed in the perceptually rigid group. Fourteen of the perceptually rigid subjects (25 per cent) would be incorrectly labeled as perceptually flexible.

On the photic stimulation task a significantly larger number of subjects fell into the perceptually rigid than into the flexible group (56 rigid, 34 flexible). Two explanations seem probable. One, our criteria was at fault. Two, the larger number of perceptually rigid subjects was a function of the photic stimulation task. Both probably are factors. Our criteria were a first attempt at such discrimination, and refinement is undoubtedly possible. However, the most important factor seems to have been the general conditions under which our sample was subjected to the photic stimulation task.

For most of these subjects it was their first trip to a neuropsychiatric hospital where the laboratory is located, and it probably was disconcerting to have several needle electrodes stuck into one's scalp for the first time. There was no structuring as to what would happen to them or what they should see during photic stimulation. Thus it is not surprising that many of the subjects were apprehensive, and reluctant and guarded about expressing themselves. Those able to express themselves freely are, perhaps, individuals who are able to adapt themselves to new and ambiguous situations.

The Rorschach, on the other hand, was administered in a more familiar situation—at the seminary attended by these students. The rigidity scores on the Rorschach were normally distributed, and the scores for both our perceptually rigid and flexible groups fall within one standard deviation of Fisher's normal group.

Comparison of Individual Rorschach Variables

To see which of the determinants were contributing most heavily to the differentiation between the two groups, a comparison of individual Rorschach variables was made. Cronbach (3) has pointed out many of the errors and pitfalls in such comparisons. Of particular importance is the positive relationship between the number of responses (*R*) and a number of clinically important determinants. Because the number of responses were significantly different between our two groups (the perceptually flexible group gave more responses than the perceptually rigid group), the chi squares for number of movement (#*M*) and number of original responses (#*O*) are reported with and without corrections for *R*. The method of correction was that suggested by Cronbach (3, p. 411). As most of the distributions were skewed, chi-square measures of significance were computed. The median of the combined groups was used as the dividing point. A contingency coefficient was computed to give an estimate of the degree of association between the two groups for each variable. These correlations, of course, are not comparable to product-moment correlations and are limited by the number of categories, but their use does not include assumptions of normality or linearity.

Table 2

Comparison of Individual Rorschach Variables

Rorschach variable	Mean scores perceptually		Chi square	p	C
	Flexible	Rigid			
R	48.3	34.5	11.53	.001	.34
#*M* (uncorrected for *R*)	4.62	2.28	20.79	.001	.44
#*M* (corrected for *R*)			7.69	.008	.28
#*O* (uncorrected for *R*)	10.2	6.1	10.76	.001	.33
#*O* (corrected for *R*)			.027	NS	—
T/R	36.1	53.9	17.80	.001	.41
T/1R	18.5	27.6	8.30	.006	.29

Of the 17 Rorschach variables for which chi squares were computed, five were significant beyond the .01 level as indicated in Table 2 (#*R*, #*M* corrected and uncorrected for *R*, #*O* uncorrected for *R*, *T/R*, *T/1R*). Those not significant were *W*%, *D*%, *Dr*%, *S*%, *F*%, #*FC*, #*CF* + *C*, *Sum C*, #*O* corrected for *R*, #*P*, *P*% and variety of content. As *O* when corrected for *R* was not significant, the difference in number of original and variety of content responses seems to be a function of difference in *R*.

These findings relate well to measures of personality rigidity used in the clinical evaluation of the Rorschach. Most Rorschach experts would agree that restriction in number of responses, few movement responses, greater time per response and delayed reaction time are major signs utilized in the determination of personality rigidity. The number of original responses is also often used as an indicator of rigidity, and even though our flexible groups gave significantly more originals, the difference failed to be significant when corrected for *R*.

The failure of *F*% and *F* + % to be significantly different in the two groups was surprising. *F* + % approached significance, *p* of .09, and was in the expected direction, but in light of the common use of *F*% and *F* + % as indicators of constriction and rigidity, we expected the difference to be larger.

Because of the differing reports (3, 5, 13) on the relationship of *M* to *R*, a correlational scattergram was plotted and as the relation-

Table 3

Comparison of Groups Using a Plus or Minus Scoring Procedure and Comparison of Individual Elements in the Fisher Scale

Individual elements	Per cent plus scores perceptually Flexible	Rigid	*CR*	*p*
All elements	33.9	43.2	2.67	.008
$F\%$	41.1	46.4	.49	—
$\#R$	3.0	19.6	2.26	.02*
$A\%$	8.8	5.4	.51	—*
$F+\%$	35.3	55.4	1.66	.06
Color	52.9	57.1	.39	—
W	23.5	33.9	1.05	—
Dd	38.2	42.9	.48	—
M	17.6	51.8	3.23	.001
Content	76.5	64.3	1.22	—
$T/1R$	23.5	35.7	1.22	—
Shading responses	52.9	62.5	.90	
$\#R+F+\%+M$	18.6	42.3	4.16	.0001

* Corrected for continuity.

ship between *M* and *R* was linear, a Pearson product-moment correlation was computed. The correlation was .50. The correlation between #*O* and #*R*, similarly computed, was .83. It is thus not surprising that our chi-square value for *M* should decrease from 20.79 to 7.69, and from 10.76 to .027 for #*O* when corrected for *R*. These findings lend support to the findings of Fiske and Baughman (5) and indicate the importance of controlling *R* when using individual Rorschach determinants for comparative purposes.

Analysis of Elements in Fisher's Rorschach Scale

The analysis of the individual Rorschach variables and inspection of the data suggested that all elements in Fisher's Rorschach scale are not contributing equally to the total score, and that a differentiating measure of Rorschach rigidity was possible by utilizing a simple plus or minus scoring procedure. In the usual procedure a weighted score is assigned as a penalty for what is considered to be excess rigidity or restrictiveness. The larger the final summation of weights, the greater is the implied rigidity. In this present analysis, instead of assigning a weighted score, only a plus was given when a penalty score was indicated. When no penalty score was indicated, a minus was given. The individual elements of the Fisher scale were similarly analyzed with the exception of card turning which was not scored because of incompleteness of data. The percentage of plus signs for each group, and the differences are given in Table 3.

As expected, the over-all difference between the two groups is highly significant. Those classified as perceptually rigid on the basis of their response to photic stimulation have a significantly larger percentage of plus scores than do those labeled perceptually flexible. The results of the analysis of individual elements are consistent with those found in the comparison of individual Rorschach variables. Human movement and number of responses were significant beyond the .05 level, and $F + \%$ at the .06 level. While the other scores were not significantly different, the differences were in the expected direction with the exception of $A\%$ and content. The use of only $\#R$, $F + \%$ and M in future research of this kind is suggested by the highly significant difference obtained when only these three elements are used for comparison.

Relation of Perceptual Rigidity—Flexibility to Intelligence and Anxiety Proneness

The results on the Rorschach suggest that the perceptually flexible group is possibly more intelligent than the perceptually rigid one, raising the possibility that the life styles in question are simply those of more or less intelligent students. To evaluate this possibility two measures of intelligence were utilized for comparison, ACE scores and a short form of the Wechsler-Bellevue (Comprehension, Vocabulary and Similarities). The two groups did not differ significantly on either the group or individual test. The mean ACE score for the rigid group was 132.8 and 135.8 for the flexible group. The mean CVS scores were 41.7 and 42.4 for the rigid and flexible groups, respectively. Thus the possibility that the style of behavior exhibited in these two perceptual situations is a function of intelligence is not supported.

As the psychological tests and psychiatric interviews of which these data are a part were used extensively to judge anxiety proneness, a comparison between the anxiety-proneness score and our measures of rigidity-flexibility was possible. This seemed to be particularly relevant since Ulett (16) has found this measure of anxiety proneness to be related to feelings of dysphoria during photic stimulation and since there are some who think of perceptual rigidity as a result of anxiety or a defense against anxiety (4).

Neither of the measures of perceptual rigidity-flexibility was significantly correlated with anxiety proneness as measured in the original study. The Pearson product-moment correlation between Fisher Rorschach rigidity score and anxiety proneness was .10, and the biserial correlation between ratings of perceptual rigidity-flexibility based on the subjective response to photic stimulation and anxiety proneness was .11.

This low relationship raises new questions as to the relationship of rigidity-flexibility to anxiety. Many readers may draw the conclusion that the anxiety measure was inadequate. This cannot now be ruled out. However, the rating of anxiety that was used was based on a number of the commonly accepted psychological measures of anxiety in use today. The Rorschach, Taylor Anxiety Scale, Saslow Screening Questionnaire, Digit Span, Mirror Drawing, Draw-A-Person, evaluation of general test behavior, and psychiatrists' ratings all were utilized in evaluating anxiety proneness.

Three other alternatives seem possible. First, the relationship between anxiety and perceptual rigidity is complex, that is, perceptual rigidity may be a successful defense against anxiety, and, therefore, some highly rigid individuals would show little or no anxiety. It is interesting that Fisher found hysterics and paranoids, two clinical groups who show the least manifest anxiety, to be the most rigid. Certainly this whole area of perceptual rigidity, supposedly the result of anxiety, is poorly defined, unclear, and needs further study.

A second alternative is that perceptual rigidity-flexibility is an "autonomous" or "conflict-free" function of the ego as discussed by Hartman (7). That is, it is a function not exclusively or necessarily developed from or in the service of drive conflicts. Klein and his co-workers (11) believe that their results seem increasingly relevant to the above formulation. They state: "It seems a worthwhile hypothesis that those aspects of perceptual process which psychoanalytic theorists currently view as 'autonomous' are themselves idiosyncratically organized within people, i.e., into 'styles of autonomous functioning' " (11, p. 3).

A third alternative suggested by Gray Walter's work (17) is that the behavior exhibited is a resultant of physiological factors not yet determined rather than being a function of anxiety. He has reported that the onset of a particularly vivid illusion or hallucination during photic stimulation was preceded or accompanied by a characteristic alteration in the electrical response of the brain. Emotional and autonomic disturbances were usually correlated with augmented response at 6–7 cycles per second from the parieto-temporal region. The use of electronically analyzed EEG records during photic stimulation may help in indicating the association of subjective sensations to electrical discharge in the brain even though no cause-and-effect relationship can be inferred at this time. Further work is in progress in this area.

Summary and Conclusion

The results of this study add new evidence to the growing belief that there are "perceptual styles" or "perceptual attitudes" which are consistent modes of response, and that the Rorschach method can be used effectively as a perceptual instrument in the study of personality. We have shown that individuals who respond in a perceptually rigid manner in one perceptual task will respond in a like manner to a different perceptual situation. The one factor which both tasks have in common is that they are ambiguous perceptual situations. The hypothesis was proposed that individuals respond in a relatively consistent manner to environmental cues, and two perceptual tasks were utilized to test this. We further hypothesized that the Rorschach method could be used as a tool for the study of perception and that Rorschach responses would reveal adaptive processes not different in kind from those appearing in other perceptually ambiguous situations. To test the hypotheses we utilized data collected on 117 freshman students from a local seminary. These students were given the Rorschach, and their subjective sensations during intermittent photic stimulation were recorded. The subjective sensations were used for comparison with Rorschach responses.

The Rorschachs were scored for rigidity according to the technique developed by Fisher. The subjective sensations were scored independently utilizing an a priori criterion for dividing the sample into two groups, perceptually rigid and perceptually flexible.

The perceptually flexible group had significantly lower Rorschach rigidity scores than did the perceptually rigid group. When the individual Rorschach determinants were examined, the perceptually flexible group was found to have given significantly more responses, more movement responses, taken less time per response, and had a significantly lower average reaction time. In contrast to findings reported to date, *F*% and *F* + % did not significantly differentiate the two groups. Neither of the two perceptual rigidity-flexibility measures was related to intelligence-test scores or to a measure of anxiety proneness. The significance of these findings is discussed and alternative hypotheses are suggested.

Received January 6, 1955.

References

1. Becker, W. C. Perceptual rigidity as measured by aniseikonic lenses. *J. abnorm. soc. Psychol.*, 1954, **49**, 419–426.
2. Bruner, J. S. Perceptual theory and the Rorschach Test. *J. Pers.*, 1948, **17**, 157–168.
3. Cronbach, L. J. Statistical methods applied to Rorschach scores. *Psychol. Bull.*, 1949, **46**, 393–429.
4. Ericksen, C. W., & Eisenstein, D. Perceptual rigidity and the Rorschach. *J. Pers.*, 1953, **21**, 386–391.
5. Fiske, D. W., & Baughman, E. E. Relationship between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.*, 1953, **48**, 25–32.
6. Fisher, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
7. Hartman, H. The mutual influences in the development of ego and id. In *Psychoanalytic study of the child*. Vol. VII. Ruth Eissler, Anna Freud, H. Hartman, & E. Kris (Eds.). New York: International Universities Press, 1952.
8. Hertz, Marguerite W. *Frequency tables for scoring responses to the Rorschach Ink Blot test.* Cleveland, Ohio: Western Reserve Univer. Press, 1952.
9. Klein, G. S. The Menninger Foundation research on perception and personality. *Bull. Menninger Clin.*, 1953, **17**, 93–99.
10. Klein, G. S., & Schlesinger, H. J. Perceptual attitudes toward instability: I. Prediction of apparent movement from Rorschach responses. *J. Pers.*, 1951, **19**, 289–302.
11. Klein, G. S., Murphy, G., Gardner, R. W., Holzman, P. S., & Schlesinger, H. J. The relation of consistencies in cognitive behavior to personality organization. 1953–54 Progress Rep. on Proj. MH–25(C). On file, National Institutes of Health, Bethesda, Md.
12. Klopfer, B., & Kelley, D. G. *The Rorschach technique*. New York: World Book, 1946.
13. Kornetsky, G., & Gerard, D. L. Effect of increasing the number of Rorschach responses on *Sum C* and *M:* A note on Fiske and Baughman's study. *J. abnorm. soc. Psychol.*, 1954, **49**, 592–593.
14. Matarazzo, Ruth, Watson, R. I., & Ulett, G. A. Relationship of Rorschach scoring categories to modes of perception induced by intermittent photic stimulation. A methodological study of perception. *J. clin. Psychol.*, 1952, **8**, 368–374.
15. Ulett, G. A., Gleser, Goldine, Winokur, G., & Lawler, Ann. The EEG and reaction to photic stimulation as an index of anxiety proneness. *EEG clin. Neurophysiol.*, 1953, **5**, 23–32.
16. Ulett, G. A., Gleser, Goldine, Starr, P., Haddock, J., Lingley, L., & Lawler, Ann. Psychiatric screening of flying personnel. Further studies toward the development of an electroencephalographic screening technique. Proj. Rep. Number 5, School of Aviat. Med., San Antonio, Texas, 1953.
17. Walter, U., & Walter, W. G. The central effect of rhythmic sensory stimulation. *EEG clin. Neurophysiol.*, 1949, **1**, 57–86.
18. Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, Karen, Meissner, Pearl, & Wapner, S. *Personality through perception*. New York: Harper, 1954.

NEW BOOKS AND TESTS

Books

Allport, G. W. *Becoming: basic considerations for a psychology of personality*. New Haven, Conn.: Yale Univer. Press, 1955. Pp. ix + 106. $2.75.

Revised from Yale's Terry Lectures, which challenge the lecturer to relate his specialty to human welfare and to a broad conception of religion, this little book is a quietly bold attempt to present a perspective in psychology on such matters as freedom, the religious sentiments, and the democratic ideal. The unifying concept is that of personality, taken as something continuously *becoming* rather than ever meaningfully identifiable on cross-sectional analysis. Personal uniqueness, selfhood, and autonomy, familiar Allportian principles, play their predictable roles in the emphasis on the person as the growing and evolving unit of psychological interest. Finally, psychology's prime task is conceived as that of determining whether "the democratic ideal is viable." Read as an impressively learned and even poetically sensitive statement of a possible relationship between psychological science and certain vital human values, *Becoming* provides a profiable experience for any psychologist concerned about either his discipline or his status as an intellectual in the republic's citizenry. Similarly, Allport's insistence on the process character of personality is useful and welcome. One may wonder, however, if such borrowings as Swedenborg's concept of the proprium is not evidence for a flirting with reification in the ideas of personality and the self, and if the avowed acceptance of Leibnitzian idealism has not led to a kind of stubborn misunderstanding of both psychoanalysis and neobehaviorism. Both seem to serve only as convenient but ghostly whipping boys. Even among those who share Allport's values and admire his humanity, there is still room to doubt that his methodological point of view does other than inhibit the kinds of understandings for which he argues so urgently and with such cultivated persuasion.—*E. J. S.*

Arieti, Silvano. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955. Pp. xxiii + 522. $6.75.

In spite of the author's modest disclaimers, *Interpretation of Schizophrenia* is a thorough, systematic treatment, and about as complete as a single volume on so vast a subject can be expected to be. Dr. Arieti proceeds logically from a historical development to psychodynamics, through a study of formal mechanisms in schizophrenia, to discussions of progressive stages of the disease, the organic manifestations, and methods of treatment. The point of view is dynamic without an underestimation of somatic factors. Of special interest to psychologists are chapters on "The Retreat from Emotions," "The Retreat of Reason," and "The Retreat from Sanity," to single out only three of the twenty-four excellent chapters. The book is well-written and reads easily; case discussions are few but rather full, although illustrative behavior samples abound; a bibliography of 280 titles is listed. It is an important reference work as well as a balanced text on schizophrenia.—*M. K.*

Blum, Lucille H., Davidson, Helen H., & Fieldsteel, Nina D. *A Rorschach workbook*. New York: International Universities Press, 1954. Pp. 166 (paper). $2.00.

This lithographed paper-bound volume introduces the student to each of Klopfer's major scoring determinants and provides an extensive series of scoring exercises for each. The "correct" answers are provided in an Appendix. A brief final chapter is devoted to administration with special attention to the inquiry as an essential prerequisite to accurate scoring. Neither the text nor the exercises are concerned with interpretation, which in the authors' opinion should follow the development of skill in scoring. The "basic bibliography on scoring" has seven titles.—*E. L. K.*

Patrick, Catharine. *What is creative thinking?* New York: Philosophical Library, 1955. Pp. xi + 210. $3.00.

This is an amalgam of well-known publications on creative thinking, with selected materials on reasoning. Included also are a chapter on age and productivity and a program for improving creative thinking. The quotations and data are organized around stages of preparation, incubation, illumination, and revision or verification. There is a comprehensive bibliography.—*A. R.*

Rose, Arnold M. (Ed.) *Mental health and mental disorder, a sociological approach*. New York: Norton, 1955. Pp. xiv + 626. $6.50.

This volume is the first of a projected series to be published under the auspices of the Society for the

---

Note.—The reviews were prepared by the Editor and the Advisory Editors, who may be identified by their initials.

Study of Social Problems, an interdisciplinary group concerned with the integration of sociological, psychological, and anthropological research and the application of these researches to human problems. It contains a collection of thirty-eight essays, case studies, research reports, and systematic analyses of the literature, planned to give an overview of the rather scattered materials in the field. Ten of the papers were prepared especially for the volume; the others have appeared previously, chiefly in sociological journals. The papers are presented in seven groups, the titles of which indicate the wide scope: problems of social psychiatry and theoretical overview, social characteristics of the mentally disordered, mental disorder in the community, sociological approaches to the study of specific disorders, social psychology of personality organization and disorganization, the sociological approach to problems marginal to mental disorder, and contributions to the understanding of mental health. The survey may be useful to psychologists. It concentrates a considerable amount of data and discussion, and most of the papers have extensive bibliographies.—*A. R.*

Royce, James E. *Personality and mental health.* Milwaukee, Wis.: Bruce, 1955. Pp. xiv + 352. $3.50.

This brief survey of modern clinical psychology is scholastic in viewpoint and "integrated throughout with a frankly Christian philosophy of life." The author planned his basic text in mental hygiene and the psychology of adjustment with reference to the problems of students. He therefore did not attempt to include references on all the material or on the various areas, but gave emphasis to classroom situations and students' needs.—*B. M. L.*

Ulett, George. *Rorschach introductory manual.* Beverly Hills, Calif.: Western Psychological Services, 1955. Pp. 45 (paper). $3.50.

This is the second edition of a manual first published in 1949 as an aid in teaching the Rorschach technique to physicians. It is an attractively printed lithographed booklet, covering the essentials of administration, scoring, and interpretation. Included also are a percentage calculation table, an ingenious multicolored "Interpretation Chart," and a table of "signs" to aid in clinical diagnosis. The approach is eclectic, and the text, while didactic, is extremely lucid. The first sentence in the text states: "The Rorschach test is a remarkably accurate means of gaining information about a patient's personality . . ."; later, in the chapter on interpretation, the author warns that "some of the Rorschach theory and interpretation given in the following pages are still open to question." There is a selected bibliography of 61 titles.—*E. L. K.*

Tests

Holzinger, Karl J., & Crowder, Norman A. *Holzinger-Crowder Uni-Factor Tests.* Grades 7–12. 2 forms. 40½ (90) min. IBM or hand scoring. Test booklets ($5.70 per 35) with manual, pp. 30; 2 answer sheets ($1.80 per 35, and $1.15 per 35); keys (50¢); specimen set (50¢). Yonkers, N. Y.: World Book Co., 1955.

The *Uni-Factor Tests* consist of nine subtests which measure four factors: verbal, spatial, numerical, and reasoning. Each factor is scored separately, and interpreted by the aid of percentile tables which give norms for each grade from 7 to 12. In addition, a Scholastic Aptitude Score may be obtained from a simple regression equation which combines the verbal, numerical, and reasoning factor scores. Alternate-form and split-half reliabilities for the separate factors compare favorably with other similar instruments. Typical intercorrelations among the factors, for single-grade groups, range from about .30 to about .50, and show that the scores, while not independent, possess considerable distinctiveness. The Manual contains unusually full data on the relationships of the factors to other concurrent measures—other intelligence tests, achievement tests, and school marks. In these data, differential values of the factors are evident. Although the ultimate merit of the test in predicting academic and vocational criteria remains to be determined, it represents one more thoughtfully constructed tool for analyzing "global intelligence" into some meaningful parts.—*L. F. S.*

Wechsler, David. *Wechsler Adult Intelligence Scale (WAIS).* Ages 16 and above. Individual test. 1 form. Test material, with 25 record forms, and manual, pp. vi + 110 ($21.00); record form only ($1.60 per 25); manual only ($2.75). New York: Psychological Corp., 1955.

The long-awaited revision of the Wechsler scale will fulfill the demands of psychologists—it is enough of a revision to remedy the major shortcomings of the 1939 Wechsler-Bellevue, and not so sweeping a change as to discard the experience gained by clinicians from the use of the older test. The eleven subtests remain the same, but all except the Digit Span have been revised in some degree. The Vocabulary subtest is wholly new; troublesome and outdated items have been removed from Information, Comprehension, and Arithmetic, and the Object Assembly has been lengthened by adding another object, an elephant. The major standardization from ages 16 to 64 was based on 1,700 examinees with equal numbers of men and women, and with a precise stratification of geographical region, urban-rural residence, race, and occupation based on the 1950 census. The manual contains appropriate data on reliabilities and subtest intercorrelations for three age segments. Corrected split-half reliabilities are .97 for the full scale and range from .66 to .96 for the subtests. The manual and record blank both warn against making clinical interpretations from profiles or from differences between subtest scores with limited reliability. Left for future research, perhaps quite properly, are the relationships between the old and new scales, and the validities of the revised test for various applications. The test itself and the data in the manual give the impression that the WAIS is a mature and sophisticated instrument, a worthy successor to the Wechsler-Bellevue for both clinical service and research.—*L. F. S.*

Wittenborn, J. Richard. *Wittenborn Psychiatric Rating Scales*. Psychiatric patients. Individual rating scale. 1 form. Rating scales ($2.75 per 25); with manual, pp. 12; specimen set (35¢). New York: Psychological Corp., 1955.

A series of 18 research articles since 1950 has reported the development and application of the Wittenborn scales. Now they are made available for general use in an attractive and convenient format. The scales are an aid to the objective description of psychiatric patients in terms of their observable symptoms. On each of 52 areas of behavior, a patient is rated by selecting the one of four statements which describes him best. The quantified ratings are turned into scores on nine factors identified by factor analytic studies. The corrected split-half reliabilities of the factor scores range from .67 to .92, with only one scale below .76; their intercorrelations are generally low. The new blank greatly simplifies the calculation of the factor scores by reducing it to the addition of digits in conveniently arranged columns and the use of a simple table. With their new convenience and availability, the scales will probably receive wider use in research. They make a valuable contribution to the knottiest problem in studies of psychopathology, the problem of the criterion.—*L. F. S.*

Books Received

Bettelheim, Bruno. *Truants from life*. Glencoe, Ill.: Free Press, 1955. Pp. xvi + 511. $6.00.

Bram, Joseph. *Language and society*. Doubleday Short Studies in Sociology, No. 8. Garden City, N. Y.: Doubleday, 1955. Pp. viii + 66 (paper). 95¢.

Chinoy, Ely. *Automobile workers and the American dream*. Garden City, N. Y.: Doubleday, 1955. Pp. xx + 139. $3.00.

Deutsch, Felix, & Murphy, William F. *The clinical interview*. Vol. II: *Therapy*. New York: International Universities Press, 1955. Pp. 335. $7.50.

Faris, Robert E. L. *Social disorganization*. (2nd Ed.) New York: Ronald, 1955. Pp. viii + 664. $5.50.

Frank, Lawrence K. *Individual development*. Doubleday Papers in Psychology, No. 14. Garden City, N. Y.: Doubleday, 1955. Pp. vii + 52 (paper). 85¢.

Freeman, Frank S. *Theory and practice of psychological testing*. (Rev. Ed.) New York: Holt, 1955. Pp. xvi + 609. $5.25.

Ghiselli, Edwin E., & Brown, Clarence W. *Personnel and industrial psychology*. (2nd Ed.) New York: McGraw-Hill, 1955. Pp. ix + 492. $5.00.

Gillespie, James M., & Allport, Gordon W. *Youth's outlook on the future*. Doubleday Papers in Psychology, No. 15. Garden City, N. Y.: Doubleday, 1955. Pp. ix + 61 (paper). 85¢.

Glover, Edward. *The technique of psycho-analysis*. New York: International Universities Press, 1955. Pp. x + 404. $7.50.

Greer, Scott A. *Social organization*. Doubleday Short Studies in Sociology, No. 9. Garden City, N. Y.: Doubleday, 1955. Pp. x + 68 (paper). 95¢.

Jewish Vocational Service. *Words for work*. Boston: Jewish Vocational Service, 1955. Pp. 140 (paper). $1.00.

Karn, Harry W., & Weitz, Joseph. *An introduction to psychology*. New York: Wiley, 1955. Pp. xii + 315. $3.90.

McCary, James L. (Ed.) *Six approaches to psychotherapy*. New York: Dryden Press, 1955. Pp. vii + 402. $3.75.

Maier, Norman R. F. *Psychology in industry*. (2nd Ed.) Boston: Houghton Mifflin, 1955. Pp. xiv + 678. $5.50.

Mayer, Jurt B. *Class and society*. Doubleday Short Studies in Sociology, No. 10. Garden City, N. Y.: Doubleday, 1955. Pp. vii + 88 (paper). 95¢.

Michaels, Joseph J. *Disorders of character*. Springfield, Ill.: Charles C Thomas, 1955. Pp. x + 148.

Munn, Norman L. *The evolution and growth of human behavior*. Boston: Houghton Mifflin, 1955. Pp. xi + 525. $6.50.

National Conference of Social Work. *Casework papers, 1954*. New York: Family Service Association of America, 1955. Pp. 150 (paper).

Oates, Wayne E. *Religious factors in mental illness*. New York: Association Press, 1955. Pp. xx + 239. $3.50.

Podolsky, Edward. (Ed.) *Management of addictions*. New York: Philosophical Library, 1955. Pp. xvii + 413. $7.50.

Ralli, Elaine P. (Ed.) *Adrenal cortex*. Transactions of the Fifth Conference, 1953. New York: Josiah Macy, Jr. Foundation, 1954. Pp. 187. $3.75.

Reitman, Francis. *Insanity, art, and culture*. New York: Philosophical Library, 1954. Pp. 111. $3.75.

Rothstein, Jerome H., & O'Connor, Thomas. *Films on the handicapped*. Washington: International Council for Exceptional Children, 1955. Pp. vi + 56 (paper). $1.00.

Thorpe, Louis P. *Child psychology and development*. New York: Ronald, 1955. Pp. v + 709. $6.00.

U. S. Department of Health, Education, and Welfare. Public Health Service. *Evaluation in mental health*. Washington: U. S. Government Printing Office, 1955. Pp. x + 292. $2.00.

Wallin, J. E. Wallace. *Education of mentally handicapped children*. New York: Harper, 1955. Pp. xiii + 485. $4.50.

Wallin, J. E. Wallace. *The odyssey of a psychologist*. Wilmington 4, Del. (311 Highland Ave., Lyndalia): Author, 1955. Pp. xvii + 243 (paper). $3.00.

JOURNAL OF CONSULTING PSYCHOLOGY

December, 1955 Vol. 19, No. 6

Contents

Comparison of Visual, Content, and Auditory Cues in Interviewing: *F. Harold Giedt* - - 407
The Stability of Interaction Chronograph Patterns in Psychiatric Interviews: *George Saslow, Joseph D. Matarazzo, and Samuel B. Guze* - - - - - - - - - - - - - 417
Changes in Some Moral Values Following Psychotherapy: *David Rosenthal* - - - - - 431
Projection, Self Evaluation, and Clinical Evaluation of Aggression: *Anthony Davids, Andrew F. Henry, Charles C. McArthur, and Leo F. McNamara* - - - - - - - - 437
Repression and the "Return of the Repressed": *John H. Flavell* - - - - - - - - - - 441
Further Validation of the Ego-Strength Scale: *Robert D. Wirt* - - - - - - - - - - 444
The Generality of the Prediction of Self Reports: *John J. Wittich* - - - - - - - - - 445
Empathy, Similarity, and Self-Satisfaction: *Howard M. Halpern* - - - - - - - - - 449
Cultural Symbolism: A Validity Study: *Emanuel Starer* - - - - - - - - - - - - 453
Temporal Acuity of Vision, Audition, and Touch in Psychogenic and Neurogenic Pathology: *Albert F. Ax and William H. Colley* - - - - - - - - - - - - - - - - - 455
Social Desirability and Q Sorts: *Allen E. Edwards* - - - - - - - - - - - - - - 462
The Reliability of Adjustment Ratings and the Length of Case Histories: *A. W. Bendig* - 463
Some Evidence on the Validity of the Sarason Test Anxiety Scale: *Barclay Martin and Bruce McGowan* - 468
Movement as a "Rhetorical Embellishment" of Human Percepts: *Leon H. Levy* - - - - 469
The Measurement of Primary Mental Abilities by the Columbia Mental Maturity Scale: *Cyril R. Mill and Charles J. Turner* - - - - - - - - - - - - - - - - - - 472
Reactions to Rorschach Cards IV and VII as a Function of Parental Availability in Childhood: *Ralph Hirschstein and Albert I. Rabin* - - - - - - - - - - - - - - - 473
Rorschach Content and Hostile Behavior: *Martin R. Gluck* - - - - - - - - - - - 475
A Statistical Comparison of the WISC and Wechsler-Bellevue, Form I: *John R. Price and Gareth D. Thorne* - 479
Diagnostic Accuracy or Diagnostic Stereotypy?: *C. H. Patterson* - - - - - - - - - 483
A Reply to Patterson: *Albert Kostlan* - 486
The Manifest Anxiety and ACE Scales and College Achievement: *Henry E. Klugh and A. W. Bendig* - 487
New Books and Tests - 488
Index of Authors of Books and Tests Reviewed - - - - - - - - - - - - - - - 491

No. 5 wanting

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 70 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 70 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 70-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. —— from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $—— for microfilm or $—— for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Extended report. The full report is prepared in the style specified by the *Publication Manual* (1), except that it may be typed with single spacing for economy in photoduplication by the ADI.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.*, 1952, **49**, 389–449.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Comparison of Visual, Content, and Auditory Cues in Interviewing[1]

F. Harold Giedt[2]
VA Hospital, Sepulveda, California

Despite the wide use and importance of the interview as a method of judging personality, it still remains largely an artistic technique rather than a scientific tool. Because of considerable "faith validity" and more modest empirical validity, the interview is most probably here to stay. Consequently, the best research approach would seem to be to determine in what way accurate or inaccurate information is obtained through interviewing, rather than to try to prove merely whether or not it works.

This study was designed, among other things, to determine which of the several forms of observations available to the clinician in the interview situation seems most related to making accurate judgments of a patient. The plan was to control experimentally the sorts of observations available to the evaluating clinician by means of interviews recorded on sound film. By this procedure the clinician could be presented with (*a*) visual cues (silent film), (*b*) verbal content (verbatim written transcript), (*c*) content plus auditory cues (sound alone), or (*d*) content plus auditory and visual cues (complete sound film).

Procedure

Interviews. Two clinicians interviewed four patients, producing eight separate interviews which were recorded on sound film. Two interviewers, a psychologist and a psychiatrist, were used to provide two different approaches to interviewing. Each interviewer was instructed at least to touch on the patient's home adjustment, work adjustment, educational achievement, and his present illness. Both were also acquainted with the personality dimensions that were later to be rated on the basis of the recorded interviews.

Four male patients were used in this study to provide some representation, albeit limited, of divergent personalities and modes of adjustment or maladjustment. Three moderately disturbed, voluntarily hospitalized, neuropsychiatric patients and one general medical patient were chosen. A brief description of each patient follows:

Patient A was a young veteran diagnosed schizophrenic reaction, paranoid type, moderate.

Patient B, a middle-aged man, was hospitalized for ulcer-like symptoms but later was diagnosed as having gastritis and a hypertrophic body in the antrum of the stomach. He represented a pleasant, near normal subject.

Patient C was a young veteran diagnosed anxiety reaction, acute, with marked tension, irritability, mild depression, and battle dreams.

Patient D was a middle-aged man who was first hospitalized for alcoholic gastritis and a possible ulcer, but was later diagnosed as schizophrenic reaction, unclassified, with a depressive reaction and chronic alcoholism.

The interviews were all filmed in one afternoon. They were conducted in an office with a one-way mirror. Camera, sound equipment, and technical personnel were located in an adjacent room out of sight of the patients. There were about twenty minutes for the patient to become accustomed to the lights and microphone equipment and for the interviewer to develop some rapport with the patient before the formal interview lasting twelve minutes began. It was the consensus of opinion of those involved that all the patients responded well to the interviews and did not appear to be particularly defensive, embarrassed, or exhibitionistic.

---

[1] From the Veterans Administration Hospital, Long Beach, California.

[2] This study is a revision of a doctoral dissertation submitted in 1951 to the Department of Psychology, University of California at Los Angeles. I should like to express my gratitude to the many professional clinicians who so kindly allowed their evaluative skills to be examined. Also I would like to thank Drs. James F. T. Bugental, Roy M. Dorcus, and John P. Seward for their aid and encouragement.

Evaluation measures. Given some sample of a person's behavior as it occurs in an interview, there are a number of possible judgments or conclusions that could be made about that person. In actual practice clinicians describe a patient's personality, predict his behavior, and make recommendations for his treatment or disposition. Such end products of the clinician's interview with a patient are usually either put in the form of written reports, or are directly translated into action of some sort with or upon the patient.

In an objective study of the interview it is important to deal with these end products in as close to their original form as possible. However, they should also be made quite explicit and subject to some form of quantification or objective analysis. Personality rating scales and prediction of patient responses to sentence completion test items were chosen as two evaluation instruments to provide some measure of the impressions gained from the interview materials.

Rating scales. In order to have rating scales that would measure those personality characteristics most commonly considered in psychological reports, case histories, and case conferences, a list of traits or dimensions was gathered from such reports and conferences, and each trait was defined. In consultations with five clinicians the original list of thirty-three variables was reduced to eleven which were considered the most necessary dimensions for describing a patient's personality. To make it possible for the rater to translate accurately his impressions of a patient into marks along a line representing variation in a personality characteristic, distances along the line were made meaningful by means of descriptive cue statements whose positions on the scales were empirically determined. The completed rating scales were constructed so that there was a separate page for each rating scale and space was provided for listing cues the judge felt he used in making the rating. Analysis of these reported cues will be reported in a later paper.

For most of the personality characteristics that clinicians are called upon to evaluate (except intelligence), there are no standardized tests that are widely accepted as criteria. The best available criterion in this case seemed to be agreement among five persons considered the best available judges of the personalities of the four patients. One judge was the chief psychiatrist at the hospital where all the patients considered were then under treatment. This psychiatrist was also one of the two interviewers in the study. Another judge was the psychologist who was the other interviewer in the study. Both made the personality ratings immediately following each interview with the four patients. Two other psychologists contributed criterion ratings on the basis of testing the patients. One administered the Rorschach and Wechsler-Bellevue test to all four patients. The other gave the Thematic Apperception Test and the Draw-A-Person test. The experimenter made another set of ratings.

The criterion, as well as all subsequent ratings, were translated into numerical values ranging from zero to nine. The five sets of criterion ratings were then statistically studied by simple analysis of variance. This was done to determine whether the differences between mean ratings of patients were significantly more than might be expected on a chance basis, as determined from differences among the criterion judges' ratings of single patients. It was felt that only if such agreement could be demonstrated could the average of the five judges' ratings on one patient be considered stable enough to be used as a criterion rating. Only five of the rated personality characteristics showed considerable agreement (p of .001) among the judges' ratings for each patient and one more showed fairly close agreement (p of .02). In addition to these six criterion ratings the Wechsler-Bellevue total IQ scores were used as the criteria for the ratings of Intelligence. Table 1 lists the seven characteristics for which acceptable criteria were established and gives the numerical value of the ratings for each patient as well as the meaning of the rating at the lowest numerical value.

Table 1

Numerical Values of Criterion Ratings for Each Patient

Personality characteristics	Patients				Meaning of low numerical values
	A	B	C	D	
Intelligence	3	6	4	2	Very superior
Anxiety	3	5	1	2	Overwhelmed with anxiety
Dependency-Independency	5	7	2	3	Wholly dependent on others
Adherence to Reality	4	6	6	3	Psychotic disregard
Direction of Feelings towards Others	6	5	5	7	Very warm feelings
Spontaneity	7	6	6	2	Wildly impulsive
Flexibility of Adjustive Techniques	7	4	7	7	Tries anything that works

Predicting patient sentence completions. In this study as in Luft's study (4), the method of controlled cross-predictions was selected as a second measure of effectiveness in judging others. This method of determining a clinician's ability to predict a patient's responses to open-ended incomplete sentences has the advantage of providing the clinician with an opportunity to predict the patient's actual behavior in a very specific, known situation, and of having the patient's own responses serve as the criteria. Its chief disadvantage and criticism may be that the behavior which is predicted has little practical significance. However, the assumption is that differences in ability to predict less crucial but still far from simple behavior, such as a patient's response to a test item, would reveal different degrees of understanding of that patient. Specifically, the method of controlled cross-predictions as used in this study required the clinician to select out of four choices the response he felt a particular patient gave in completing an incomplete sentence. The four choices were the actual responses given by the four patients interviewed in this study. An example of one such prediction situation follows:

Item 12. Most of all I want (stimulus portion of the sentence)

—to succeed (patient C's response).

—peace of mind (patient D's response).

—travel and a good looking wife (patient A's response).

—a good job (patient B's response).

There were a total of thirty-five incomplete sentences for which the clinicians had to select the actual responses given by the particular patient under consideration and the number of correct predictions constituted his accuracy of prediction score. This measure had a split-half reliability of .60 with the Spearman-Brown Correction.

Presenting the interview material. A number of factors had to be considered in designing the exact manner in which the interviews were to be presented. There were (*a*) the four modes of presenting the interview material, (*b*) the four patients to be judged, (*c*) the two interviews with each patient, and (*d*) the order in which any condition was presented. In order to study the effects these various independent variables had upon the dependent variables, the two measures of accuracy in making personality judgments, a modified greco-latin square design was employed. All factors were varied and controlled so that no condition had any priority in the

Table 2

Design for Presenting Interview Material

Evaluating clinicians	Order of presentation			
	1	2	3	4
Group I	Aa1	Bb1	Cc2	Dd2
Group II	Bc2	Ad1	Da2	Cb1
Group III	Cd1	Dc2	Ab1	Ba2
Group IV	Db2	Ca2	Bd1	Ac1
Group V	Aa2	Bb2	Cc1	Dd1
Group VI	Bc1	Ad2	Da1	Cb2
Group VII	Cd2	Dc1	Ab2	Ba1
Group VIII	Db1	Ca1	Bd2	Ac2

Note.—Capital letters (A, B, C, D) represent the four patients. Small letters (a, b, c, d) represent the four modes of presenting the interview material. Numbers (1, 2) represent the two interviewers. Each group was composed of two psychiatrists, two social workers, and two psychologists.

order of presentation and no combination of conditions was favored, as can be seen in Table 2.

The evaluating clinicians. A most important group of participants in this study were the psychiatrists, social workers, and psychologists, who graciously volunteered to observe the different sorts of interview material and then attempted to evaluate the patients who were interviewed. Altogether there were forty-eight such evaluating clinicians, with sixteen from each of the three specialties. Each group indicated in Table 2 was composed of a pair from each of the three professional groups with one of the pair being an experienced clinician and the other someone in training.

Results

Personality ratings. Each clinician's rating of a particular patient on a particular trait was compared to the criterion rating for that patient and trait, and the amount of deviation determined. A total deviation or rating-error score for each clinician's set of ratings on each patient was obtained by summing the deviations from the criterion ratings for the seven personality dimensions. To insure that the deviations for all seven dimensions would contribute equal weight to the total rating-error score, the variances of the seven distributions of deviations were equated, and where necessary, the deviations were multiplied by a constant before they were summed to form the total rating-error score. A low score indicates an accurate series of ratings.

In order to get an idea of the statistical and psychological meaning of this total error score, the intercorrelations between the seven distributions of deviations from the criterion ratings were determined. The average correlation was .03 with only two positive and one negative coefficients of the twenty-one intercorrelations significantly above zero at the one per cent level. This finding suggested two possible explanations. The different ratings are in fact quite unrelated, and one would not expect accuracy in judging one trait to carry over to another trait. The other possibility was that the trait ratings are very unreliable measures and deviation from or approximation of the criterion is governed largely by chance.

To determine which of these explanations is the more applicable the intercorrelations between the actual ratings, disregarding the criterion ratings, were found for the seven rating scales. The average intercorrelation was only .16. Eleven positive and two negative *r*'s of the matrix of twenty-one intercorrelations were significantly larger than zero at the one per cent level. These correlations suggest that the actual trait ratings were, aside from some halo effect, relatively independent ratings and lend support to the first explanation. Although no measure of repeat reliability of the ratings was found, an approximation can be obtained in the intercorrelation of the scales, *Direction of Feeling towards Others* and *Social Impression,* which was .60. One scale was designed to measure the patient's general friendliness and the other the judges' friendliness towards that patient, dimensions which are probably very closely related. Thus it would seem that the repeat reliability would be greater than .60. The interjudge reliability ranges from an *eta* of .55 to .68 indicating a fair degree of agreement between the judges in their ratings.

It therefore seems reasonable to conclude there is sufficient reliability in the ratings and independence in the scale dimensions to justify combining the separate deviations from the criterion ratings into a total error score as a measure of the accuracy of each rater's judgment of a particular patient.

When the data were tabulated, no even moderately significant over-all difference was found to exist between the accuracy of ratings based on interviews conducted by Interviewer No. 1 (psychiatrist) and Interviewer No. 2 (psychologist). However, objective differences in the way in which the two interviewers operated and their effects on the patients' responses has been demonstrated by Bugental (1) using the written protocols from this study.

Since there was no demonstrable difference in the accuracy of ratings due to different interviewers, and in order to simplify further statistical treatment of the data, all the ratings on the same patient based on a particular mode of presenting the interview material were combined. This meant, in effect, super-imposing the bottom half of Table 2 on the top half.

Analysis of variance. Data derived from the four effects or variables listed below were analyzed:

1. Order in which a patient was observed and rated.
2. Sequence or pattern in which different patients and modes of presenting the interviews were observed and rated. This effect also contains any differences between the four combined groups of clinicians as each group followed one of the four sequences of presentation.
3. Patient being observed and rated.
4. Mode of presenting the interview material.

From Table 3 it is apparent that the factor of the patient being rated had the greatest effect on the accuracy of the ratings. Patients A and B, who were most accurately judged, have most of their criterion ratings at or near the theoretical mean and have the largest number of ratings on the favorable ends of the scales.

The mode of presenting the interview material appears to have some influence over the accuracy of the ratings. Actually, the only statistically reliable differences (p of .05 or better) between the different modes are between the silent film and the other three modes and between the written protocols and complete sound film.

There was a significant order or practice effect, in that a clinician's ratings tended to improve in accuracy from the first to the third time he filled out the rating scale, but then dropped on the last time. The drop might have been due to the fact that most clinicians rushed through the last rating, as they were free to leave when they finished it.

Because of the partial confounding of main effects with interactions in the greco-latin square, it was impossible to measure directly the interaction between specific variables. Fig. 1 suggests considerable interaction between the mode of presentation and the patient being rated. The relative contributions of order, sequence, and interaction, however, cannot be computed. Another indication of interaction effects is that the variance estimate based on the square residual, a rough indicator of interaction and error, approaches significance (p of .05).

From Figure 1 it is clear that, particularly when only visual cues were presented, the four patients' mean accuracy ratings differ widely. When only the written transcript of the interview was presented the mean accuracy of ratings for all four patients showed much more agreement than for any of the other three modes of presentation.

Predictions of patient responses. The accuracy of predicting a patient's actual sentence completion response was not appreciably affected by which of the two clinicians conducted the interview, so, as before, the scores were combined for the same patients and conditions of presentation.

It can be seen from Figure 2 that when the clinicians saw only the silent movie of the interview their predictions of the patient's responses were poorer than chance. In other

Table 3

Analysis of Variance on the Accuracy of Ratings

Source of variance estimate	Sums of squares	Degrees of freedom	Variance estimate	F ratio
Order	205.7	3	67.57	5.86***
Sequence	75.2	3	25.07	2.12
Mode of presentation	188.6	3	62.87	5.38**
Patient	274.3	3	91.43	7.81***
Square residual	133.0	3	44.33	3.87*
Within cells (error)	1,949.0	176	11.07	
Total	2,825.8	191		

* Significant at .05 level of probability.
** Significant at .01 level of probability.
*** Significant at .001 level of probability.

Bureau Ednl. Psy. Research
DAVID H. ... NING COLLEGE
Dated ....
No. .. ..

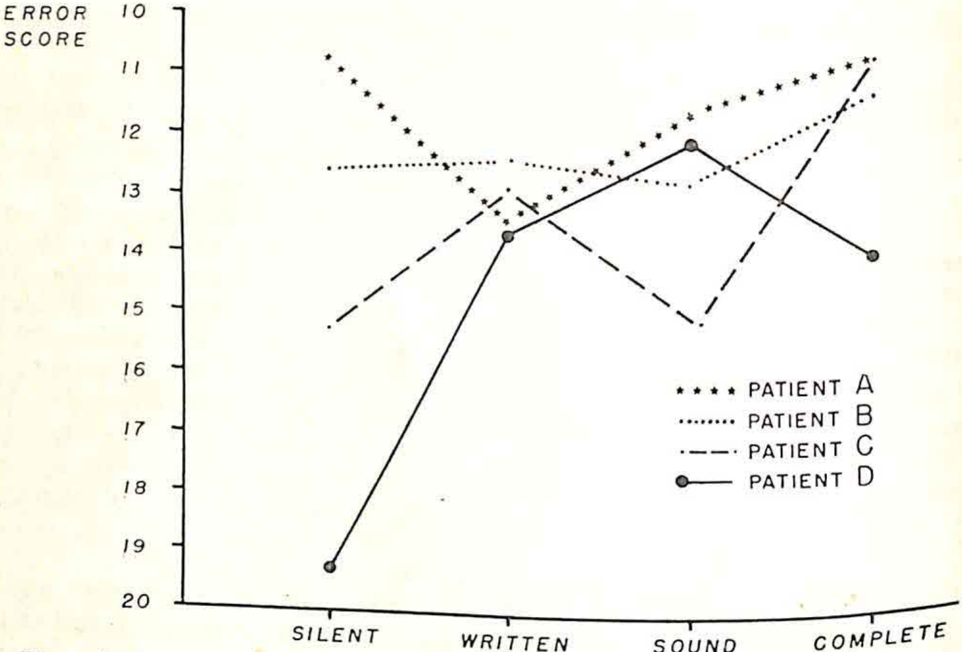

Fig. 1. Mean accuracy of personality ratings of four patients under different conditions of observation. Lower error scores indicate greater accuracy.

words, under such conditions the clinician was more apt to be wrong than right in his predictions. For all the other conditions the predictions were considerably better than chance.

To determine how many predictions one could expect on a chance basis alone, the empirical probabilities of designating each of the four choices for all the thirty-five completed sentences were determined. The prob-

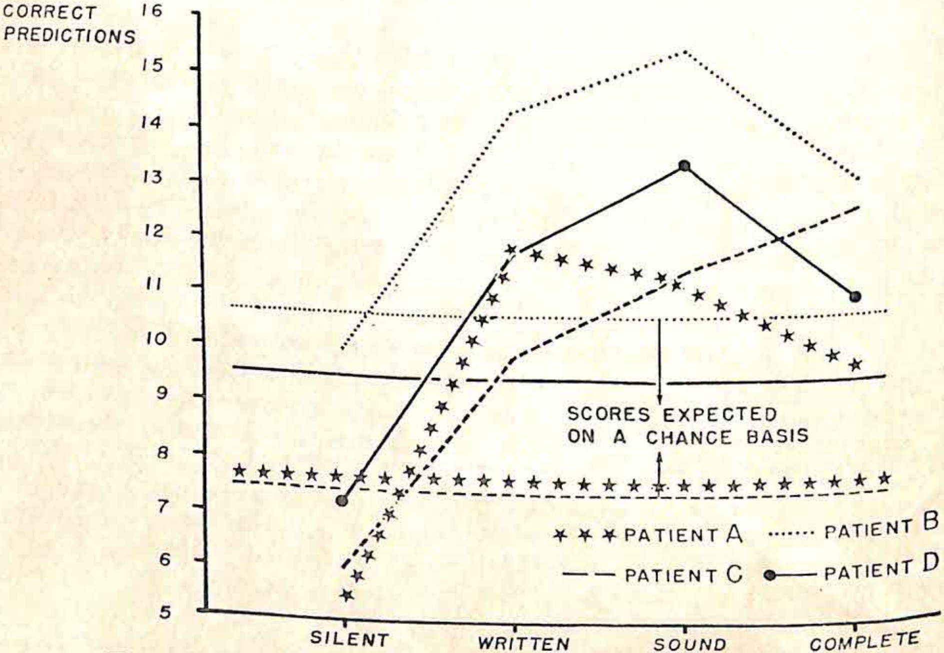

Fig. 2. Mean accuracy for predicting responses of four patients under different conditions of observation.

Table 4

Analysis of Variance on the Accuracy of Predictions of Patients' Responses

Source of variance estimate	Sums of squares	Degrees of freedom	Variance estimate	F ratio
Order	26.9	3	8.97	—
Sequence	32.8	3	10.93	1.05
Mode of presentation	974.9	3	324.97	31.3***
Patient	383.2	3	127.7	12.3***
Square residual	171.7	3	57.23	5.52**
Within cells (error)	1,827.2	176	10.38	
Total	3,416.7	191		

** Significant at .01 level of probability.
*** Significant at .001 level of probability.

abilities for the thirty-five correct responses for each patient were summed and are represented by horizontal lines in Figure 2.

In the analysis of variance done on the prediction scores, the same four variables or effects were considered as in the case of the ratings. Table 4 gives the results of the analysis of variance. Clearly the mode of presenting the interview material has a most significant effect on the number of correct predictions the clinicians were able to make. Actually the only statistically reliable differences were between the silent film presentation and each of the other three modes.

As in the case of the ratings, it was also found that the particular patient being considered influenced the accuracy of predictions. The order and sequence of presentations produced no significant effects. However, the square residual was found to yield a significant F ratio suggesting some possible interaction effect. In comparison with the effect of different modes of presentation, this square residual is so much smaller that little doubt is cast on the significance of the single effect of the mode of presentation. Figure 2 seems to indicate further no appreciable interaction such as was the case in Figure 1.

Relationship between predictions and ratings. The rating-error scores and the number of correct predictions were correlated and the signs of the coefficients reversed to indicate the degree to which the accuracy of personality rating was related to the accuracy of predicting patient responses. The resulting correlation coefficients are contained in Table 5.

Besides showing very little relationship between the accuracy of the two evaluation instruments, there is even a suggestion that under the condition of sound only the two instruments are measuring inversely related phenomena.

Discussion

Several quite certain, several speculative, and several practical interpretations can be made of the results of this study. When trying to rate a patient's personality or predict his behavior, it helps considerably to know what he says about himself in an interview situation, whether one reads or hears what he had to say. Where the predictions of sentence completions were made from just seeing a patient, the results in all four cases were poorer than could be expected on a chance basis. The increment in accuracy was considerable between the silent film presentation and the three other presentations (Figures 1, 2) which all had verbal content cues in common. One

Table 5

Relationship Between Predictions and Ratings under Different Conditions

Mode of interview presentation	r	$\frac{r}{\sigma_r}$	Number of measures correlated
Silent film	.28	1.92	48
Written transcript	.22	1.51	48
Sound recording	−.31	−2.12*	48
Complete sound film	.03	.20	48
All conditions combined	.14	2.00*	192

* Significant at .05 level of probability.

might expect such results on a common sense basis except for the discovery that as a group the judges actually did worse than if they had used a chance system or predicted an "average patient." It may well have been that judges were misled by dress, appearance, gestures, or facial expressions.

Another definite finding was that some patients can be much more accurately predicted or rated than others. However, although the "normal" subject's responses were best predicted he was not the easiest to rate. Also with the ratings some patients were more accurately rated on the basis of one type of observation but were quite poorly rated in other situations. Thus there do not seem to be clear and consistent differences in the accuracy with which different patients can be judged that hold up for different media of observation and different forms of appraisal. There is some evidence that patients whose sentence completion responses were more average-sounding, and frequently selected on a purely chance basis, were more accurately predicted and that those patients whose criterion ratings were mostly near the average range were more accurately rated. Gage (3) has specifically studied these types of effects.

Of considerable interest is the lack of any clear and consistent relation between judges' ability to make accurate ratings and their ability to make correct predictions. Also one wonders about the trend toward continued improvement with increased cues for the ratings as compared to the drop in accuracy in the complete sound film presentation for the predictions. Since they were not appreciably related, which of the two measures is the "truest" measure of ability to judge others, and consequently the most valid measure to use in comparing the value of different interview cue material? Quite clearly the two judging processes are different. According to Wallins' classification (6) the ratings would be considered analytical-type judgments, involving some degree of abstraction and the predictions would be nonanalytical or empathic-type judgments. Taft (5) suggests there is no necessary relationship between these two methods of assessing ability to judge others, which was well demonstrated in this study. Whereas Taft points out that personalities and other characteristics of judges are associated with success on either type of measure, results from this study suggest that the cues available to judges also have a differential value for the two types of judgments.

Another explanation exists for the differences shown in Figures 1 and 2, where the ratings show gradual improvement with addition of cues while the predictions improve up to the sound-only condition, but then drop in accuracy with the complete sound film. In the former condition the judges, with each added bit of observation data, were approximating the unlimited access to the patient which the criterion judges had. Thus their accuracy improves as the experimental situation approximates the criterion situation, with possible biases based on a patient's appearance operating both on criterion and experimental situation judges. Where the criterion was based on the patient's actual free response to an open-ended incomplete sentence, opportunity to observe the patient appears, if anything, to have had a slightly detrimental effect on accuracy. Although the results of this study are only suggestive, this difference has important implications for prediction studies. For instance, if a criterion such as job success is determined largely by supervisors with certain consistent biases or errors in judgment, prediction of "successful employees" should somehow include or take account of these biases or errors in the criterion so as to produce the most accurate predictions. However, if the criterion is some sort of performance or achievement measure little influenced by consistently biased superiors, prediction of success would of course be best accomplished by measures free of bias and conducive to objective appraisal. From this point of view, neither of the techniques for determining ability to judge others can be considered more "true" or basic, but both are appropriate representatives of different kinds of judgments we make about others.

Because the differences in accuracy of ratings and predictions between the three conditions containing content cues were not statistically significant, the safest conclusion would be that, in general, with this type of information-gathering, diagnostic type of interview, there are no appreciable differences introduced by the different sorts of cues available to the judge as long as the content of the interview was available to him. The interesting trend towards poorer predictions with the addition of visual cues to content and auditory cues should be subjected to further study with a greater number of patients to be observed, and limitation of the study to only the two conditions, so as to reduce the amount of variance introduced from other less relevant sources. If one accepts the conclusion that the three conditions produce essentially similar results, a number of practical recommendations can be made. For financial or other considerations it is frequently possible to obtain only verbatim transcripts or sound recordings of an interview for use in research or other purposes. The question has been raised as to whether such material is an adequate representation of the original session, or whether a complete sound film was necessary. It would appear on the basis of this study that written protocols or sound recordings are quite adequate, and perhaps even superior where the purpose is to gain predictive knowledge of a person.

Although this study was not done with the employment or selection interview as the focus, some of the conclusions from this study suggest variations in selection interviewing that might be found to be useful and advantageous. Often at considerable expense and inconvenience to both applicant and employer, a personal interview is arranged. If the results obtained in this study hold for the assessment of employees, then it might be possible for the applicant to be interviewed by telephone or for him to submit either a sound recording or a written protocol of an interview conducted by some recognized local interviewer. The employer could then have this "canned" interview appraised without actually having to see the applicant. Fay and Middleton (2) report better results in selecting sales personnel by a similar technique. Whether such a procedure would emotionally satisfy an employer who has become accustomed to "seeing anyone I hire" is another question.

The noticeable reduction in variation of mean rating error scores in the written transcript condition shown in Figure 1, invites some explanation. In the other three conditions the appearance of some expressive behavior of the patient is observable and may tend either to aid or mislead the clinicians. With the written transcript there are only the bare words used by the patient, which evidently have more nearly the same significance or cue value to the clinicians. Specific instances of the helpful and confusing effects of these noncontent cues will appear in another paper. Also because it was impossible to control whether a clinician reread parts of the written transcript, or to know how well each person could hear the recorded spoken words, variation in accuracy due to misunderstanding was doubtlessly reduced with the written protocols.

Because this study was based on recorded interviews and the judges were limited to observing someone else interview patients, the results cannot be immediately generalized to all "live" interviewing. They do, however, suggest that interviewers should be careful to avoid being misled by a patient's appearance, should critically consider how they use expressive cues, and should continue to rely heavily on what a patient says.

Summary and Conclusions

Interviews with four patients were recorded on sound film to study the sorts of observations that aid the clinician in making accurate judgments of a patient's personality and predictions of his verbal behavior. The interviews were presented as silent films, written transcripts, sound recordings, or complete sound films to forty-eight psychiatrists, social workers, and psychologists, who made ratings of the personality characteristics and predicted the responses to incomplete sentences of the patients who were interviewed.

The following conclusions can be drawn from this study:

1. There was marked and significantly greater accuracy of personality ratings when content cues were included, as in the written transcript, sound recording, and complete sound film, as compared to the silent film. Although this difference was the only clearly significant one, there was a general trend

with considerable variation from case to case for the ratings to improve in accuracy as more and more cues were added.

2. Markedly and significantly better predictions were made from written transcripts, sound recordings, and complete sound films than from silent films which actually led to predictions which were worse than could be expected on a chance basis. An interesting difference from the trend towards gradual improvement with additional cues found with the ratings was seen with the predictions where somewhat poorer predictions were made in the complete sound film condition. The suggestion that visual cues or a patient's appearance may impair predictions should be subjected to further study.

3. As might well be expected, significant differences were found in how well each patient was either rated or his responses predicted. In the case of the predictions, the level of predictive accuracy varied for different patients but the pattern for each patient over the four modes of presentation remained quite similar. However, with the ratings some patients were more accurately rated in one condition and less in another, suggesting considerable interaction between the effects of the patient being rated and the type of observation.

4. No appreciable relationship was found between accuracy of personality ratings and correct predictions, which probably can be accounted for on the basis that the two evaluation techniques may tap quite different aspects of personality, and also are based on somewhat different sorts of cues.

5. No definitely superior medium was found for recording diagnostic-type interviews for studies of judgment of others so that any choice between written protocols, sound recordings, and complete sound films can be made on the basis of practical considerations, or factors not covered by this study.

Received April 19, 1955.

References

1. Bugental, J. F. T. Explicit analysis of topical concurrence in diagnostic interviewing. *J. clin. Psychol.*, 1953, **9**, 3–9.
2. Fay, P. J., & Middleton, W. C. Relationship between sales ability and ratings of the transcribed voices of salesmen. *J. appl. Psychol.*, 1942, **26**, 499–509.
3. Gage, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
4. Luft, J. Implicit hypotheses and clinical predictions. *J. abnorm. soc. Psychol.*, 1950, **45**, 756–759.
5. Taft, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.
6. Wallin, P. The prediction of individual behavior from case studies. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 48, 181–250.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

The Stability of Interaction Chronograph Patterns in Psychiatric Interviews[1]

George Saslow, Joseph D. Matarazzo
Massachusetts General Hospital[2]

and Samuel B. Guze
Washington University School of Medicine

The objective assessment of personality is an old problem in psychological research. During the past fifty years many approaches to this problem have been undertaken, which have included questionnaires, paper and pencil inventories, projective tests such as the Rorschach and TAT, sentence completion tests, and others. The results with these various instruments, while sometimes positive, have been, in the minds of many investigators, less than encouraging. These investigators have felt that to attempt to assess fixed or unchanging intrapsychic traits such as levels of anxiety, the presence or absence of "defenses" of various kinds, etc., by such instruments was to commit the "organism error" described by MacKinnon (14, p. 43). This error involves the search for stable or invariant personality characteristics which will be present independent of the stimulus situation or "field" in which one is attempting to observe or record these traits. For this reason, many research workers, as well as many practicing clinicians, use the equally unreliable (used here in the sense of unstandardized) "interview" as their instrument of assessment. One reason given for this is that the interview is flexible and allows one to explore unique areas of behavior for each individual. Also, the interview presents a "dynamic" stimulus situation so that every individual is given an opportunity to manifest his unique, and presumably learned, social or interpersonal behavior patterns (18).

The notoriously low inter-interviewer agreement on assessment of personality patterns of functioning for given series of subjects (12, 15) has made the interview subject to criticism also. It is important, therefore, if the interview is to be an instrument of research, to ask what could be achieved if the interview were standardized. Chapple (1–6), working primarily in applied anthropology, has been developing just such a standardized interview during the past several years. A review of this development and of the instrument, the interaction chronograph, used for making measurements of the interaction between interviewer and interviewee during the standardized interview, is now in preparation and may be consulted for a more thorough description than is possible below.

Essentially, the interaction chronograph is a device which allows an observer to record in time units with a high degree of precision the behavioral interaction[3] of two individuals in terms of some ten or more variables. These variables, definitions of which are given in Table 1, are objectively recorded by a series of electrically controlled counters which are connected to two keys, one for the interviewer, the other for the subject. Each key is depressed whenever the designated individual is

---

[1] This investigation was supported by a research grant (M-735) from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

[2] The data of this study were collected while all three authors were at the Washington University School of Medicine.

[3] Records include only such behavior as number of utterances, number of interruptions, their durations, etc., and not the content of the verbalizations.

Table 1

Definitions of Ten Interaction Chronograph Variables for Which Reliability Was Determined

1. *A's Units*: The designation "A" is used rather than "Pt.'s units" since interviewees other than patients are often recorded. This counter provides a *frequency* count of the actions of the interviewee. Each time he is active, and the observer depresses his key, the unit counter adds one unit. Thus a cumulative record is made of the number of times the individual was active during a period designated by the signal counter. In view of the fact that different people have different units of action during comparable interviews or time periods, Chapple divides each of the following variables (2–9) by the number of A's units and thus obtains mean values which make possible comparisons between individuals.
2. *Tempo*: The present machine does not record interviewee actions and silences separately (these, however, can be computed readily when desired). Rather, this counter records the *duration* of each action plus its following inaction in a single measure. Thus it provides an index of how often one starts to act; the duration from one action to his next action; or his "tempo" of starts.
3. *Activity*: This, too, is a *duration* measure derived from an individual's actions and silences. However, unlike Tempo which is a measure of actions *plus* silences, the activity counter adds actions and subtracts silences. The cumulative record indicates, therefore, for any interview sequence, how much more active a person was than he was silent. Thus, the activity counter is an index of a subject's "energy level," states Chapple.
4. *A's Adjustment*: This counter, unlike Tempo and Activity, does not operate all the time the machine is on. It records only *durations* during which the person *interrupted* B (the other person) or *failed to respond* to B following the latter's last action; both of which are failures of adjustment by A. The counter adds the durations of interruptions and subtracts from these the durations of failures to respond. Thus a positive "adjustment" score indicates that on the average A interrupted longer than he was silent. A negative score indicates the converse. One would expect a type of "manic" subject to earn a positive score while some "depressed" individuals should earn negative adjustment scores.
5. *B's Adjustment*: The other individual's adjustment (i.e., the adjustment of the interviewer, usually). For a description substitute "B" for "A" in (4) above.
6. *Initiative*: This is a *frequency* counter which records only after a period in which both A and B were silent (both keys up). It provides a measure of the relative frequency with which one person, A, takes the initiative as against the other person, B. It adds just one each time A takes the initiative and subtracts one each time B takes the initiative. A person who initiates more often than he has to be initiated to earns a plus score whereas a negative score indicates the converse.
7. *Dominance*: This is a *frequency* counter which records only during an interruption (double action, with both keys down). It provides a measure of the relative frequency with which one person, A, out-talks or out-acts the other person, B, when there has been an interruption. The Dominance counter adds one if A dominates B in a double action, or subtracts one if B dominates A. A positive score indicates that A was more dominant than B in these exchanges, and a negative score the converse.
8. *A's Synchronization*: This is a *frequency* counter which records the *number of times* A has interrupted B or failed to respond to B, i.e., the number of times A failed to synchronize with B. The earlier described counter, A's Adjustment, records the relative duration of these events; the present counter their frequency.
9. *B's Synchronization*: The other individual's synchronization (i.e., the synchronization of the interviewer, usually). For a description substitute "B" for "A" in (8) above.
10. *B's Units*: This counter provides a *frequency* count of the actions of the other person (usually the interviewer). It adds one unit each time the observer depresses B's key during an interactional exchange.

talking, nodding, gesturing, or in other ways communicating (interacting) with the second person. Values for these variables are cumulative and can be abstracted from the printed record of the total interview with little difficulty. Some of these variables may seem unusually arbitrary, since they represent algebraic sums of two variables rather than individual measures of each of these variables. Apparently Chapple, in developing his interaction theory of personality, has found these derived variables more useful than the first-order variables from which they were obtained. In addition to containing individual counters for each variable, the interaction chronograph has a "signal" counter which functions as a marker to record the start of different periods of the interview.

Chapple (2) and Goldman-Eisler in England (7, 8, 9, 10), working independently have found that the use of this instrument for the objective assessment of personality patterns is made more difficult by the differences in "interaction patterns" of different interviewers. Furthermore, any one interviewer may use with any given patient only a small portion of the entire range of interactional behavior available to him, as shown in a recent study of a diagnostic interviewer by Saslow *et al.* (17). Thus, it has been demonstrated that the differences in inter- and intra-interviewer interaction patterns have a subtle but very marked effect on the interviewees' interaction patterns when these are carefully and objectively recorded. These experimental results help to define some of the uncontrolled variance which in the past has made the interview, including the interviewer's behavior, a less than adequate research tool. Experimenters have attempted to use the interviewer as an *independent* variable in their efforts to measure various interviewee characteristics (the *dependent* variables). But experience has indicated, and the results of Chapple, Goldman-Eisler, and Saslow *et al.* confirm, that the "interviewer" is not an objective scientific instrument and that, in fact, interviewers are themselves dependent, in the sense of unstandardized, variables.

The effect of differences between interviewers is very convincingly demonstrated in a recent study by Raines and Rohrer (15). Although using an approach (psychiatrists filled out rating scales and gave psychiatric diagnoses following a screening interview) which was different from that used by both Goldman-Eisler and Chapple, Raines and Rohrer conclude:

> The results of this study support a "projection" hypothesis to account for the variance between psychiatrists observed here; that is, that one of the major variables responsible for variation between psychiatrists in impressions of patients is the differing personalities of the psychiatrists themselves (**15**, p. 732).

It is little wonder then, for example, that the science of psychotherapy has made so little headway when its effects have been "evaluated" by various interviewers (16).

Chapple has developed a *standardized interview* in order to facilitate the process of making the interviewer (and the conduct of the interview) an independent variable, and thereby to permit comparison of results in different settings obtained by different interviewers, or results obtained by the same interviewer at various times. Chapple's standardized method involves certain "rules" for the interviewer to follow in his own interviewing behavior and in the over-all conduct of the interview itself. The standardized interview is divided into five periods; with periods 1, 3, and 5 as free give-and-take periods, and period 2 (silence) and 4 (interruption) as *stress* phases of the interview. The characteristics of this standardized interview are shown in Table 2, whereas the "rules" governing the interviewer's standardized behavior are given in Table 3.

Thus Chapple's standardized interview provides a means for eliciting a sample of patients' behaviors (dependent variables) in a complex but miniature, molar, interpersonal situation, the characteristics of which to a cer-

Table 2

Characteristics of the Standardized Interview

Period	Type of interviewing	Duration of period Fixed duration		Variable duration
I	Free	10 minutes		
II	Stress (Silence)			12 failures to respond or 15 minutes, whichever is shorter
III	Free	5 minutes		
IV	Stress (Interruption)			12 interruptions or 15 minutes, whichever is shorter
V	Free	5 minutes		
Total		20 minutes	plus	a maximum of 30 more minutes

Table 3

Standardized Interviewer's Behavior

Rules for Interviewer

Periods 1 to 5 (all periods):

a. Interviewer introduces each period by a 5-second utterance (following his signal to the observer).
b. All interviewing must be *nondirective*. No direct questions, no probing or depth interviewing. Interviewer can reflect, ask for clarification, ask for more information, introduce a new topic area, etc. In general, interviewer's comments should be nonchallenging and open-ended and related to the patient's past comments or to some new, general topic.
c. All interactions must be verbal only, or verbal and gestural at the same time; i.e., interviewer cannot use head nods and other gestures alone. This rule simplifies the observer's task.
d. All of interviewer's utterances must be of approximately *5-seconds* duration.
e. After patient finishes a comment or other interaction, interviewer *must* respond in less than *1 second*, except as otherwise noted in Period 2.
f. Each time patient interrupts interviewer, the latter must continue to talk for 2 more seconds. This rule insures more explicit definition of a patient's ascendance-submission pattern than would be possible if interviewer "submitted" immediately.

Periods 1, 3, and 5:

a. Interviewer must never interrupt patient.
b. If after interviewer makes a comment patient does not respond, interviewer must wait *15 seconds* and then speak again for 5 seconds.

Period 2 only:

a. Interviewer must "fail to respond" to last interaction of patient a total of 12 times (or for 15 minutes, whichever is shorter).
b. After interviewer has been silent for 15 seconds (and patient has not taken initiative) interviewer makes another 5-second comment.

Period 4 only:

a. Each time patient acts, interviewer must interrupt patient for 5 seconds for a total of 12 times.
b. Interviewer's interruption should begin about 3 seconds after patient has begun his interaction.
c. After having interrupted patient, if the patient continues through the interruption (does not submit), interviewer will not interrupt again until patient has finished his utterance, i.e., interviewer will interrupt patient only once during each utterance of the latter if patient does not "yield."
d. The period is ended after 12 interruptions or 15 minutes of attempting to obtain these.

tain degree are objective and predefined (independent variables). Such controlled "field" situations would seem to hold promise of improved assessment of personality.

The aim of the present study was to investigate the consistency of scorable interaction patterns elicited from patients interviewed independently by two different interviewers in a standardized manner. Thus the question was asked, "Will patients behave in the same way in independent interviews with two interviewers, if the latter employ similar interviews and interviewing behaviors?" Put differently, the question was asked: Is it possible to obtain inter-interviewer reliabilities of patient interaction patterns of magnitudes higher than those obtained by Raines and Rohrer (and numerous other investigators) when the behavior patterns of the interviewers are more controlled and less subject to individual, uncontrolled variation? It was recognized at the outset that the problem of the reliability of the measuring instrument was *confounded* with the question of the variance or invariance of the interaction patterns being measured by this instrument. It was reasoned that high reliability cofficients on the various variables would tend to indicate *both* stability of interaction patterns of patient-subjects and also the reliability of the instrument of measurement; whereas low or insignificant correlations would not permit one to decide whether the unreliability was in one of these factors, or the other, or both.

Procedure

After they practiced the interview technique described in Tables 2 and 3 in some 40 interviews each, the objective interaction scores of the interviewers indicated that they were approaching the prescribed patterns with a high degree of accuracy. This was achieved without the interviewers feeling unduly inhibited in their responses to the patients. At this point the present study was initiated. The design called for twenty patients, each to be interviewed independently by the two interviewers, one of whom was a young internist experienced in interviewing, and the other an older psychiatrist. The *Ss* were outpatients sent to the psychiatric clinic of a large Midwestern clinic for screening and assignment to individual psychotherapists. They were new patients, white, 11 men and 9 women, ranging in age from approximately 18 to 55. The presenting problems were typical of the population of this outpatient clinic, and consisted of cases of anxiety reaction, hysteria, depression, schizophrenia, obsessive-compulsive neurosis, duodenal ulcer, and possible chronic brain syndrome.

The design was randomized to control order effects and each interviewer thus interviewed every other patient first; an AB order with the first patient and a BA order with the second, etc. In this way each interviewer came first in the sequence for ten patients and second for the remaining ten. Statistical evaluation of these "interviewer-order" effects on patients' behavior was therefore possible.

Ash trays were removed from the interviewer's desk so that patient smoking would be discouraged. No mention was made of the experiment by the interviewer, and the experiment began when the interviewer opened the interview with such a statment as "My name is Dr. ——. Can you tell me how you happened to come to the clinic at this time?" The interviewer pressed a concealed button for a light signal to the observer. The observer was seated in the next room on the other side of a large one-way window (5 ft. × 2 ft.) and recorded the interaction by pressing designated buttons for the interviewer and the patient. A high-fidelity microphone was hung from the ceiling and connected to an amplifier in the observer's room. The side effects of the recorded voices in the observer's room were controlled by piping the voices to the observer through earphones. Throughout the experiment, sound recording and visual observation of the interactions were excellent. Verbatim recording of the interview exchanges was made by means of recently devised highly sensitive microphones, connected directly to an Audograph machine in the observer's room. The Audograph discs were transcribed for subsequent studies (content analyses, etc.).

The second interviewer was always in another building when the first interviewer was conducting his interview, thereby insuring independence; both began with zero knowledge about the patient. After finishing his interview, the first interviewer would say, "Mr(s). ——, I'd like another doctor in our clinic to talk with you now, and if you will wait here a minute or two I will go to get him." Upon his arrival, the second interviewer signalled the observer and the second half of the experiment began.

Subjects were used in the order in which they were sent to this particular clinic. However, due to power failure in the hospital on one occasion, and to total incoherence on the part of one patient, two of the consecutive planned experiments were lost.

Results

From Table 2 it is seen that twenty minutes of the interview are fixed and that, depending upon the subject's behavior in periods 2 and 4, thirty more minutes are possible to complete the standardized interview. Table 4 contains an analysis of the actual mean length of the twenty pairs of interviews. It can be seen that the standardized interview takes, on the average, about 33 minutes to conduct, and that both doctors had essentially equal interview times. Furthermore, the order of the interview, comparing all 20 *first* interviews or all 20 *second* interviews—independent of which doctor was first or second—yielded no differences in mean length of interview

Table 4

Analysis of Mean Length of Total Interview for 20 Patients Interviewed Twice

Analysis	Mean	Range
Length of all 40 interviews	32.8 minutes	25.7 to 50.3
a. Length of all first interviews ($N = 20$)	32.9 minutes	26.9 to 41.4
b. Length of all second interviews ($N = 20$)	32.8 minutes	25.7 to 50.3
c. Length of all first doctor's interviews ($N = 20$)	33.5 minutes	26.9 to 50.3
d. Length of all second doctor's interviews ($N = 20$)	32.2 minutes	25.7 to 41.2

Table 5

Means, Standard Deviations, Ranges and Coefficients of Correlation Across Total Interview for Each of the Major I–C Variables

Variable	Dr. 1	Dr. 2	r	p
1. Mean Pt.'s Units	72.20	69.85	.747	.01
SD	25.55	20.90		
Range	25 to 127	29 to 112		
2. Mean Pt.'s Action*	59.85	56.25	.956	.01
SD	56.90	54.60		
Range†	13 to 243	17 to 257		
3. Mean Pt.'s Silence	9 20	8.30	.759	.01
SD	3.11	2.41		
Range†	5 to 17	3 to 14		
4. Mean Tempo	57.20	52.85	.908	.01
SD	36.87	27.26		
Range†	24 to 159	26 to 142		
5. Mean Activity	39.55	36.20	.930	.01
SD	40.49	30.01		
Range†	−2 to 150	2 to 131		
6. Mean Pt.'s Adjustment	−1.53	−1.28	.802	.01
SD	1.33	1.17		
Range	0 to −5	0 to −4		
7. Mean Dr.'s Adjustment	−1.83	−1.53	.737	.01
SD	1.17	.80		
Range	−3.57 to +1.00	−3.03 to +.14		
8. Mean Pt.'s Synchron.	.84	.84	.726	.01
SD	.07	.06		
Range	.63 to .99	.76 to .98		
9. Mean Dr.'s Units	62.79	59.74	.772	.01
SD	22.78	18.28		
Range	21 to 109	27 to 93		

* Values for variables 2 through 7 are given in hundredths of a minute. To convert to seconds, multiply the given value by 0.6.
† Rounded to nearest whole number.

when subjected to statistical analysis. It is of interest to remember that, relatively speaking, no restrictions were placed on the lengths of two of the periods (i.e., patients could vary in their behavior in these periods such that the lengths of these periods could range from, say, one minute to fifteen minutes). Table 4 indicates that even with this freedom to vary, the patients behaved similarly in regard to this variable with the two interviewers. The mean lengths in minutes for the two doctors, respectively, were 8.41 and 8.44 in period 2 ($r = .631$), and 3.68 and 2.26 in period 4 ($r = .555$). These differences were not significant by F test, while the values of r were significant at the .01 level of confidence. These first observations thus point to the existence of stability in the interaction patterns of patients in a standardized interview.

Table 5 presents, for the 20 patients, the means, standard deviations, and ranges (for the total interview) of nine of the interaction chronograph variables. Values are given for each doctor, with Pearson coefficients of correlation shown in the second column from the last. The results show a striking *stability* in interaction variables from first interview to second interview, for each of the twenty patients.

The stability in patients' interaction patterns (and corollary reliability of measurement by the interaction chronograph) is demonstrated in Table 5 in several ways: (*a*) *relative* reliability between the behaviors manifested by each of the 20 patients with the two doctors, and (*b*) *absolute* reliability of these same interaction patterns.

Relative reliability is inferred from the values of the Pearson product-moment coefficients of correlation shown in the second column from the last. These coefficients range from .726 to .956 and are all significant at the

.01 level of probability. They indicate that, for the interview as a whole (summing all five periods), each of these 9 interaction chronograph variables has a marked stability, i.e., that two interviewers will elicit these patterns from each of the 20 patients in amounts which are relatively similar for any one of these individuals.

However, since one can get high values of Pearson r for any one variable when 2 interviewers elicit *different amounts* of this variable from the same patient, providing the *relative* values of these amounts are maintained from patient to patient,[4] these high values for r are in themselves not sufficient to demonstrate reliability in the sense of stability (or invariance) in patient characteristics.

To demonstrate the latter, one needs to examine the values for the means, *SD*'s, and ranges for any one variable across the two interviewers. Examination of Table 5 indicates an equally striking *absolute* reliability for these 9 variables for the 20 patients with the two interviewers. Thus, for example, despite very marked individual differences in the interaction rate of the 20 patients (pt.'s units), as seen by the fact that patients varied from 25 to 127 units with Doctor 1, and 29 to 112 with Doctor 2, the average number of units for these 20 patients was 72 with Doctor 1 and almost 70 with Doctor 2. That is, despite the fact that, during an interview of 33 minutes' average length, one patient interacted 25 times (less than once a minute) and another patient 127 times (4 times per minute) with Doctor 1, and similarly with Doctor 2, any *single* individual seemed to maintain his *own* interaction rate independent of which doctor was interviewing him. Thus one individual interacted 25 times with the first doctor and 29 times with the second, while another subject interacted 112 and 129 times, respectively, with the two interviewers. This marked similarity in values of the means, as well as the equally striking similarity in standard deviations and ranges for all the variables, would tend to indicate, when added to the high values of the Pearson r's, that these interaction variables reflect stable and invariant personality characteristics under these standardized conditions in the subjects studied. Before leaving Table 5 several other findings should be pointed out. The first is the finding that how active ("action") a patient is on the average, is a more stable or invariant characteristic from interviewer to interviewer ($r = .956$) than is any other variable measured. Average length of silence, on the other hand, while fairly stable is less so ($r = .759$). Tempo and activity, which are algebraic summations of action and silence, are very stable (.908 and .930) and yield values similar to those reported by Chapple on 12 subjects (2). Chapple gave no reliability figures for the other variables.

It is of interest to note also that the two interviewers interacted with the patients in similar ways. From Tables 2 and 3 it can be seen that no attempt is made to standardize *how often* the interviewer should interact with the patient. Except for somewhat defining his number of units in periods 2 and 4 (thus controlling his behavior in 12.8 minutes of the total 32.8 minutes) he is free to follow the patient's initiative in regard to this variable during the three remaining free periods (20 minutes). Despite this freedom the two doctors had similar average numbers of Dr.'s units of interaction (63 and 60, with an r of .772) and similar "adjustments" (algebraic sum of durations of interruptions and failures to respond).

Table 5 contains values for only 9 of the 10 variables defined in Table 1. Values for Dr.'s synchronization were not included because inspection revealed a very limited range in the scores for this variable—at or very near 1.00 throughout—making statistical analysis of little usefulness. The value for r for Dr.'s synchronization was $-.058$, while the values for the means, *SD*'s, and ranges were .98, .029, and .94 to 1.04 for Doctor 1; and .96, .050, and .84 to 1.02 for Doctor 2, respectively. The values for this variable are an internal check, and indicate that the two interviewers did in fact follow the standardized rules very closely and, except in periods 2 and 4, which

---

[4] One could get a perfect correlation (1.00) on pt.'s units, for example, if the second interviewer *always* got double or some other multiple of the number of units from the same patients; i.e., if from 3 pts. one doctor got 30, 40, and 50 units, and the second obtained 60, 80, and 100, the value of r would be 1.00.

Table 6

Coefficients of Correlation for Two Interviewers for Each of the Major I–C Variables in Each of the Individual Periods of the Interview

Period	Pt.'s units	Pt.'s action	Pt.'s silence	Tempo	Activity	Pt.'s adjustment	Dr.'s adjustment	Pt.'s synchronization	Dr.'s units
I	.726	.685	.845	.665	.593	.809	.476	.203	.715
II	.554	.817	.899	.802	.832	.136	.864	.836	.784
III	.365	.485	.121	.496	.471	.334	.212	.024	.436
IV	.508	.553	.849	.435	.609	.811	.696	.625	.591
V	.517	.735	.479	.736	.735	.820	.771	.006	.432
Total	.747	.956	.759	.908	.930	.802	.737	.726	.772

Note.—For $r = .444$, $p < .05$; for $r = .561$, $p < .01$.

will be discussed later, they neither interrupted nor failed to respond, therefore earning scores near 1.00 on this variable during the various periods. Due to the nature of the construction of the present form of the interaction chronograph, the very brief pauses of the doctor which occur in give-and-take interaction before he responds to the patient are recorded as if they were failures to respond. Thus an interviewer who adequately follows the standardized interview behavior will score one failure to respond for each of his units, and thus will earn a score of 1.00 instead of 0.00 on synchronization.

A word regarding the shape of the distribution of scores for each of these variables is in order. All variables yielded normal, or bell-shaped distributions, except action and silence (and their derivatives—tempo and activity). These latter yielded **J**-shaped distributions with skewing toward the high end. The recent findings of Norton (quoted in Lindquist) would indicate that little precision was lost in not normalizing the individual scores prior to statistical analysis (13, pp. 78–86). If normalizing is felt to be necessary the various transformation methods used by Goldman-Eisler with similar data are applicable (7, 8).

The remarkably stable results shown in Table 5 represent values determined on the basis of the *total* interview. However, in view of the fact that the interview (and interviewer behavior) described here is standardized for each of its five subparts, it is of interest to examine the consistency or stability of patient interactional behavior in each of these separate periods. Coefficients of correlation are given in Table 6, while means and standard deviations are shown in Table 7, for the analysis by periods.

As one would expect, and statistical theory can describe (the Spearman-Brown phenomenon), individual samples (periods) are in general less reliable than the total sample (overall interview). Thus patients' interactional behaviors during the 10-minute first period, or 5-minute fifth period, for example, are less reliable from interviewer to interviewer than are the values based on the total 32.8 minute interview.[5] Values for the total interview are reproduced from Table 5 for comparison.

There are several striking findings shown in Table 6. Of the 45 correlations shown for the 9 variables in the 5 periods, 27 are significant at the .01 level and 10 reached the .05 level of statistical significance. Thus, only 8 period-variable combinations failed to reach statistical reliability. Three of these 8 are contributed by the patient's synchronization variable—an artifact which results from the very limited range of scores found for this variable and described earlier for doctor's synchronization.

Reading across variables for each of the 5 periods, one is struck by the consistently low reliabilities found in period 3. Thus, while periods 1, 2, 4, and 5 have, respectively, 7, 8, 7, and 5 r's which reach the .01 level and

---

[5] An example analogous to the situation here described is the low test-retest reliability of some of the Wechsler-Bellevue subtests (.60), while the reliability of the Full Scale IQ (derived by a weighted adding of these subparts) is remarkably high, with an r of .97 (19, p. 13).

Table 7

Means and Standard Deviations for Two Interviewers for Each of the Major I–C Variables in Each Period

Period		Pt.'s units		Pt.'s action		Pt.'s silence		Pt.'s tempo	
		Dr. 1	Dr. 2	Dr. 1	Dr. 2	Dr. 1	Dr. 2	Dr. 1	Dr. 2
I	M	19.55	18.35	74.45	65.25	7.40	8.65	82.25	74.30
	SD	10.10	9.20	68.02	45.92	2.44	2.50	66.90	44.30
II	M	12.85	12.35	60.90	69.40	12.70	11.55	73.90	81.10
	SD	3.02	2.24	59.40	81.60	8.24	6.33	56.57	79.56
III	M	9.35	10.75	67.75	79.25	10.70	9.30	78.80	88.90
	SD	5.05	5.70	44.70	127.76	4.52	2.72	45.35	127.80
IV	M	17.55	15.20	21.00	8.85	6.45	5.80	27.75	15.10
	SD	6.22	3.54	49.74	3.47	3.10	1.78	49.55	3.75
V	M	12.90	13.20	75.90	59.30	8.40	7.40	85.50	67.05
	SD	7.14	6.10	124.12	67.82	2.60	1.80	122.82	66.71
Total	M	72.20	69.85	59.85	56.25	9.20	8.30	57.20	52.85
	SD	25.55	20.90	56.90	54.60	3.11	2.41	36.87	27.26

Period		Pt.'s Activity		Pt.'s Adjustment		Pt.'s Synchronization		Dr.'s Adjustment		Dr.'s Units	
		Dr. 1	Dr. 2	Dr. 1	Dr. 2	Dr 1.	Dr. 2	Dr. 1	Dr. 2	Dr. 1	Dr. 2
I	M	67.15	56.85	−1.17	−1.81	.94	.96	− .57	− .47	19.05	17.79
	SD	69.35	47.55	1.28	1.47	.07	.05	.26	.20	9.67	8.73
II	M	48.30	57.94	−1.11	− .99	.25	.25	−9.90	−8.95	3.21	3.10
	SD	63.10	84.00	2.13	1.34	.18	.15	5.54	4.12	2.31	2.20
III	M	57.20	69.95	−2.16	−1.54	1.02	1.00	−1.06	− .74	9.68	10.63
	SD	44.65	127.75	2.31	1.86	.04	.08	1.13	.59	5.19	5.56
IV	M	14.50	3.10	−1.53	− .99	.98	.99	.91	1.28	17.84	15.21
	SD	50.25	4.45	1.80	2.41	.07	.06	1.08	.70	5.30	2.04
V	M	67.65	51.85	−1.50	−. 95	.96	.95	−.81	− .61	12.74	13.00
	SD	125.10	68.75	1.09	1.20	.06	.08	.54	.27	6.28	5.03
Total	M	39.55	36.20	−1.53	−1.28	.84	.84	−1.83	−1.53	62.79	59.74
	SD	40.49	30.01	1.33	1.17	.07	.06	1.17	.80	22.78	18.28

1, 0, 2, and 3 which reach the .05 level, period 3 has *no* values at the .01 level and only 4 at the .05 level.

Since period 3 is a free give-and-take period similar to periods 1 and 5, the explanation for this discrepancy must be sought elsewhere than in the structure of the period itself. A tentative hypothesis is that the stress of period 2 (silence) is such as to make unpredictable from interviewer to interviewer the immediately following behavior of the same patient. Put differently, a group of 20 patients will react to silence stress (as seen in their immediately following behavior, period 3) differently, depending upon the doctor who is causing the stress. A still further description of this same phenomenon is: interaction patterns of patients seem relatively invariant from doctor to doctor under all conditions explored other than that involving silence stress.

The reason for this finding is not apparent at present. It may reflect, among other things, a true patient-instability in this regard relative to other patient characteristics, or it may reflect actual differences in doctors' silence behavior in period 2. Slight evidence for the latter was demonstrated by an analysis of the effect of "order" of interviewer upon patient interaction patterns. By use of a covariance design utilizing two conditions (an AB order for 10 *S*s and a BA order for the other 10 *S*s), values for the over-all interview for each of the variables were compared. Thus it was possible to determine, for example, if Dr. 1 elicited more units (etc.) from the same patients relative to Dr. 2, depending upon whether he, Dr. 1, interviewed these patients first or second. Despite individual differences in the means for these variables for the 10 patients interviewed under condition I, relative to the values for the 10 *different* patients interviewed under condition II, there were no differences in the means (M_1 versus M_2, and M_3 versus M_4) between the two doctors, for eight of the nine variables. Thus, for these variables, it made no difference which doctor interviewed first or which interviewed second.

This was not the case with the patient's "average duration of silence" variable, however. For this variable it was shown that in the second series of 10 patients, Doctor 2 coming first had the significant effect ($F = 4.66$, $p = .05$) of eliciting more silence in these 10 patients when they were seen immediately following this experience by Doctor 1. On the other hand, Doctor 1 did not produce this effect on the 10 patients he saw first as reflected in their "silence" behavior pattern with Doctor 2, who saw them immediately afterward. The values of M_1 and M_2 were 7.9 and 9.1, while the values for M_3 and M_4 were 9.3 and 8.7. Thus Doctor 2 had the effect of changing the patients' average duration of silence with Doctor 1 from 7.9 to 8.7 when he, Doctor 2, came first, whereas the silence of these patients with him was the same whether he came first (9.3) or second (9.1). If this is the variable which accounts for the marked effect (instability) shown in period 3 and discussed earlier, it is certainly a very subtle factor and could only be discovered by an instrument of measurement as sensitive as the interaction chronograph. Further experiments with the silence variable are therefore indicated.

It is of interest to note that spontaneous comments from three "normal" subjects interviewed by the standard interview method indicated that they found the "silence treatment" more stressful than the "interruptions." These three subjects, like a dozen or so psychologists and psychiatrists who have observed one or more of these interviews through a one-way screen, were relatively unaware of the interruption stress, i.e., they did not perceive the change in interviewer behavior in period 4. The silence stress was apparent, however. Possibly there are cultural factors involved here. Each of us has a long history of reinforcement in regard to being interrupted by others. However, in our culture, it is very unusual for anyone to fail to respond to us. Thus, silence stress is an unusual experience and is therefore possibly more disrupting than interruption. This speculation would appear to have some relevance in view of Hebb's theoretical analysis of the search by various organisms for an optimal level of excitement or stimulation from their environment (11, pp. 551–552).

Another point worth noting in Table 6 is the *r* of .136 for patient's adjustment in period 2. Because of the nature of period 2 and the method of measurement, this period-variable *does not* reflect the similarity of the patient's adjustment with the two doctors. Rather it reflects the correlation between the behaviors of the two doctors. This counter records the "durations of the patient's interruptions" and "durations of the patient's failures to respond to the doctor's last utterance." However, in period 2 the doctor's standardized behavior involves not responding to the patient in an effort to assess the latter's level of initiative. Therefore the patient's adjustment counter can record few if any failures of the patient to respond to the doctor and still fewer interruptions of him since he is not responding. The *r* of .136 is an artifact and reflects the very restricted range derived from the relatively few times the two doctors restimulated the patients during this period with the subsequent failure to respond on the part of the patients.

A truer reflection of the reliability across two doctors for the *patients'* adjustment under the stress of silence will be found in the Dr.'s Adjustment column in period 2. This counter, recording how long each doctor failed to respond to the patient before the latter took the initiative, thereby stopping this counter, yields a clearer picture of the stability of a patient's adjustment to silence. The *r* of .864 indicates that patients were remarkably similar in their adjustment to the interviewer's silence in the two interviews.

Finally a word should be said about the relatively higher *r*'s found in period 2 (most of them in the .80's) as compared to the *r*'s of period 4 (with five *r*'s below .60). Why should patients show more invariance *during* silence than during interruption? (Note this is not incompatible with their subsequent reflection of this "greater" silence stress in period 3 as discussed earlier.) The answer is not found in the present data and therefore more study is required. However, from a purely empirical point of view it can be pointed out that period 2 is a much "slower" period, containing on the average little two-way interaction, as such. Accordingly it is an easy period for the observer to record. Thus one would expect high reliabilities. Period 4, on the other hand, typically involves a "fast pace" with both patient and doctor talking at the same time. The observer has his most difficult time during this period. Hence, a possible explanation for the lower values of *r*. In order to check this point, as well as the reliability of observation during the total interview, we have designed a study in which single interviews will be recorded *simultaneously* and independently by two observers using two interaction chronographs.

Chapple has stated that the behavior of patients during the two stress periods yields important information regarding individual personality make-up. Accordingly, he has provided two measures of this behavior. The first is the *Initiative* shown by patients during the interviewer's silences of period 2, while the second is the *Dominance* pattern of the patient during the interruption stress of period 4.

Initiative is defined as the frequency with which a patient can respond or interact again when his partner fails to respond to him. It is measured in the interaction chronograph by a frequency counter which *adds* one each time the subject initiates an action following a period during which both people are silent and *subtracts* one each time the interviewer starts an action (because the subject failed to do so for 15 seconds). This figure will be negative if the interviewer is forced to start more actions than the subject, and positive if the subject shows more initiative by starting more actions than the interviewer. In order to control for differences in number of units during this period, the net count described above is divided by the total number of units during period 2.

Dominance is a measure of an individual's relative ability to "talk others down" in double actions. It is a reflection of one's ascendance-submission pattern. Measurement of this variable is made only during the fourth (interruption) period. The obtained score is the difference in the number of times one person out-talks or out-acts (dominates) the other when both have been talking or acting at the same time. If the subject dominates the interviewer the greater number of times, the score is positive; if he submits more frequently to the interviewer's interruptions, the score is negative.

Results for these two variables are shown in Table 8. Here one finds a striking stability in the initiative variable ($r = .805$) and a lower but still significant reliability ($r = .470$) for the dominance pattern. The earlier comments regarding possible observer error in period 4 are again pertinent here and stress the need for the mentioned observer-reliability study. The mean values for initiative and dominance in Table 8 indicate that patients

Table 8

Means, Standard Deviations, Ranges, and Coefficients of Correlation for Two Interviewers for Initiative (Period II) and Dominance (Period IV)

Variable	Dr. 1	Dr. 2	*r*
1. Mean initiative (Period II)	.75	.77	.805
SD	.18	.19	
Range	.17 to .94	.23 to 1.00	
2. Mean Dominance (Period IV)	−.32	−.42	.470
SD	.31	.30	
Range	−.71 to +.83	−.92 to +.29	

show a fairly high rate of initiative under silence and tend, on the average, to submit more than they dominate when interrupted by an interviewer. There are, however, very marked individual differences in these interaction patterns, as can be seen in the *SD*'s as well as in the ranges of variation.

Further effects of the silence and interruption stress can be seen in Table 7. Here it is shown that, relative to their own values in the other four periods, patients have a longer length of "silence" in period 2 and a considerably shorter length of "action" in period 4. The longer average patient silence in period 2 is of interest since it indicates that patients' interaction patterns, while relatively invariant on any one day from interviewer to interviewer, when these latter are following a standardized pattern, are nevertheless sensitive to changes in the stimulus situation (intra-interviewer's behavior). Thus in period 2, when the interviewer deviated from his period 1 pattern of responding immediately at the end of the patient's utterance (from silences of less than 1 second, to 15 seconds), the effect was to produce a concurrent increase in the patient's own duration of silence (from 7.40 to 12.70). It was subtle observations such as these which led Chapple to standardize the interviewer's behavior.

At this point a word might be said about the implication of this latter finding in regard to the question of stability or instability of interaction patterns. The results discussed earlier indicate that when the behavior of interviewers is standardized to some degree one gets highly consistent patterns from subjects. However, this invariance is for comparable samples of inter-interviewer behaviors and not for intra-interviewer changes. Thus, changes in the interviewer's behavior from periods 1 to 2 are associated with changes in interviewee reaction patterns, with the new patterns having generally high reliability. It would appear, therefore, that there is invariance (reliability) for *comparable* stimulus conditions and lack of invariance (changeability) for *changing* stimulus conditions. Without the latter one would be faced with an inflexibility of interaction behavior which would make the study (and eventual predictability) of such patterns of limited usefulness in psychology.

The question can be raised here as to whether the obtained stability in interaction patterns from interviewer to interviewer for any one interviewee was due to the latter's repeating the "same story" to the second interviewer, or some similar artifact. This was probably not so since verbatim records of the interviews show that there were often major differences in content between the first and second interviews. Military experiences might be a prominent feature of the first interview, yet hardly alluded to in the second; resented desertion by a spouse might dominate one or more subsections of the first interview, and not be mentioned in the second, etc. The records show also that psychotherapeutic changes could occur in the first interview and be reported in the second, prefaced by a clearly identifying remark such as, "Until I talked to the other doctor today, I didn't realize that . . ." etc. Some patients were obviously less emotional in the second interview than in the first. These differences between the two consecutive interviews in content, attitude, insight, emotionality, etc., will be examined in detail by a variety of methods of analysis of the verbatim materials. However, in spite of these clinically obvious differences in patients in the two interviews, there was still a marked stability in interaction patterns for any one patient.

Some postinterview comments made either to the interviewers or to doctors in other clinics, such as "I liked you more than the other doctor—I didn't like him at all," "I was upset by those doctors—I revealed things and said things I hadn't thought of for years," etc., emphasize that (*a*) the standardized interview is compatible with certain patient reactions common to any initial psychiatric interview, (*b*) the interactional pattern of the interviewer is not mentioned by the patient who is presumably not fully aware of it, and (*c*) the possibility exists of utilizing the standardized interview as a systematic observational device that can be interpolated a number of times at different points in a prolonged psychotherapeutic or other interactional relationship, without "accommodation" or practice effects.

Discussion

The results of this study would tend to indicate that considerable stability of patient interaction patterns is demonstrable if there is some control of the stimulus situation. Control of the stimulus has been overlooked all too often in psychological and psychiatric research with the interview. Our experience has indicated that the interviewer may be a sufficiently controlled or independent variable without at the same time introducing unnatural or stereotyped interviewing behavior. Neither the *S*s themselves nor uninitiated, though experienced, professional interviewers who observed the interview indicated by their comments that the standardized interview, when conducted by an experienced interviewer, is very different from other psychiatric interviews.

However, the demonstration of reliability under the conditions here investigated does not exhaust the questions concerning reliability. The obtained reliability indicates a stability in patient personality patterns with two interviewers on the *same* day. We are currently conducting a reliability study in which a single interviewer interviews the same patient on two occasions one week apart. Thus the completed study reported here involved same-day reliability across two interviewers, while the study now under way is a test-retest reliability study over time, with a single interviewer. Finally, since the high reliabilities reported here for our first experiment are unusual in psychological research involving complex variables such as one finds in the interview, we have felt it necessary to cross validate our findings and are currently doing this in a replication utilizing an identical design.

There are several other aspects of the reliability problem which should be pointed out. The first of these is the reliability of the observer's recording of the interview. He is an integral part of the instrument and the final results will be no more reliable than the reliability or accuracy of his input. We plan soon to check this aspect of the reliability problem by having two observers record the same interview, independently, across a series of patients.

Also to be examined is the reliability of the scorer who abstracts the objective scores from the numbers printed by the counters. The interaction chronograph yields cumulative scores on the several variables and thus scoring, which involves primarily simple arithmetic skills, is a reasonably objective procedure. For the study here described a second person followed Chapple's manual of instructions and scored independently the interaction chronograph records of ten of the interviews selected at random. The results indicated perfect agreement between two scorers on 96 per cent of the 600 individual final scores involved. The magnitudes of the errors involved in the remaining 4 per cent were very minimal; they were of the order of 1 unit in whole number scores and .06 unit in those variables involving hundredths of a unit. Thus it would appear that scorer reliability presents no problem in these observations.

A question which is of considerable theoretical as well as practical interest is the relation between content of verbalizations and the interaction patterns of individuals during the interview. Chapple has taken the position that these patterns are relatively invariant and are negligibly related to the topics one is discussing (4, p. 33). We plan to investigate this interesting question in subsequent research. We plan also to examine the possibility of differences in interaction patterns in various diagnostic groups. There is reason to believe that schizophrenic patients may have patterns different from those of hysterics, or of manic-depressive patients, etc. So far, only a few patients have been examined with regard to the possibility that interaction chronograph variables may reliably discriminate one category from another (5). The interaction chronograph may permit a refinement in differential diagnosis by its potential for objective or operational definitions of patient characteristics.

Finally, the question of *validity* of the interaction chronograph scores must be considered. Establishing the reliability of an instrument is only the first step. Consideration of the meaning of the obtained scores is, of course, a necessary second step. The question of validity is a complex one and will be approached in a number of different ways (predictive, content, concurrent, and construct validity). The study described above involving different patient groups is one of these approaches. However, there are various others. For example, one can ask, What are the behavior correlates in real-life situations of these measured interaction chronograph variables? Other approaches might involve use of the method in assessing the effects of psychotherapy in time; the immediate test-retest changes in interaction patterns following the administration of controlled doses of drugs which induce affective changes that can be reliably communicated by verbal report; the influence of status and role of the interacting participants upon the measurements, etc. Finally, an approach which may have considerable importance is the possibility of developing techniques for planned behavior therapy; i.e., will having the interviewer increase the length of his utterances with depressed patients increase the duration of utterances of the latter?, etc.

Summary

The present study was concerned with two problems: an investigation of the reliability of the interaction chronograph as an instrument of research and the corollary problem of the variance or invariance of interaction patterns of patients during a standardized psychiatric interview. Twenty outpatients were interviewed independently by two different interviewers under previously defined and practiced conditions. The highly significant correlations obtained indicate reliability of the instrument and a marked invariance or stability in patient interaction patterns when the stimulus conditions are relatively standardized (inter-interviewer consistency) and a flexibility in these interaction patterns when these stimulus conditions are changed in a predefined manner (intra-interviewer variance). A number of methodological questions were discussed and brief descriptions of research now under way were given.

Received August 5, 1955.
Early publication.

References

1. Chapple, E. D. Quantitative analysis of the interaction of individuals. *Proc. Nat. Acad. Sci.*, 1939, **25**, 58–67.
2. Chapple, E. D. The interaction chronograph; its evolution and present application. *Personnel*, 1949, **25**, 295–307.
3. Chapple, E. D. The standard experimental (stress) interview as used in interaction chronograph investigations. *Hum. Org.*, 1953, **12** (2), 23–32.
4. Chapple, E. D., with Arensberg, C. M. Measuring human relations: an introduction to the study of the interaction of individuals. *Genet. Psychol. Monogr.*, 1940, **22**, 3–147.
5. Chapple, E. D., & Lindemann, E. Clinical implications of measurements of interaction rates in psychiatric interviews. *Appl. Anthrop.*, 1942, **1** (2), 1–11.
6. Chapple, E. D., & Gordon, Donald G., Jr. A method for evaluating supervisory personnel. *Harvard Business Rev.*, 1946, **24**, 197–214.
7. Goldman-Eisler, Frieda. The measurement of time sequences in conversational behavior. *Brit. J. Psychol.*, 1951, **42**, 355–362.
8. Goldman-Eisler, Frieda. Individual differences between interviewers and their effect on interviewees' conversational behavior. *J. ment. Sci.*, 1952, **98**, 660–670.
9. Goldman-Eisler, Frieda. A study of individual differences and of interaction in the behavior of some aspects of language in interviews. *J. ment. Sci.*, 1954, **100**, 177–197.
10. Goldman-Eisler, Frieda. On the variability of the speed of talking and its relation to the length of utterances in conversation. *Brit. J. Psychol.*, 1954, **45**, 94–107.
11. Hebb, D. O., & Thompson, W. R. The social significance of animal studies. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 532–561.
12. Kelly, E. L. Theory and techniques of assessment. *Annu. Rev. Psychol.*, 1954, **5**, 281–310.
13. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
14. MacKinnon, D. W. The structure of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald Press, 1944. Pp. 3–48.
15. Raines, G. N., & Rohrer, J. H. The operational matrix of psychiatric practice: I. Consistency and variability in interview impressions of different psychiatrists. *Amer. J. Psychiat.*, 1955, **11**, 721–733.
16. Saslow, G. Psychotherapy. *Annu. Rev. Psychol.*, 1954, **5**, 311–336.
17. Saslow, G., Goodrich, D. W., & Stein, M. Study of therapist behavior in diagnostic interviews by means of the interaction chronograph. Paper read at Amer. Psychiat. Ass., Los Angeles, May, 1953.
18. Sullivan, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.
19. Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corp., 1955.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Changes in Some Moral Values Following Psychotherapy

David Rosenthal

Johns Hopkins University, School of Medicine

One problem facing the researcher in psychotherapy is that of accounting for the successes achieved by therapists who differ widely in both their theories of neurosis and its treatment. Eysenck (6) has questioned the validity of all these reported successes. Others, less skeptical, have considered the question of what these therapists do or have in common, reasoning that this must be the salient therapeutic force. Fiedler (7) in an investigation of three orientations to psychotherapy, the Freudian, Adlerian, and Rogerian, concluded that it was the quality of the relationship which was common to successful therapists of all three schools.

However, the question is still open with respect to what transpires, as a result of this relationship being offered, that makes for a successful outcome. Therapists generally offer the explanation that the relationship permits the unlearning of some responses and the learning of others to occur, but they do not fully agree as to what constitutes the critical learning. Rogers (11) is coming to believe that it is a learning to accept fully and without fear the "positive feelings of another." Mowrer (10) feels that the patient has to unlearn problem-solving behavior which has been used to prevent certain sign-learning from occurring. Shoben (13) sees therapy as remedying defects in social learning, the patient learning to react without anxiety to human interactions which involve impulses that have become associated with punishment. Dollard and Miller (4) emphasize the fact that the relationship permits previously inhibited fear responses to occur so that they may be extinguished.

Mowrer (9) counterposes his own position to that of the psychoanalysts, which is essentially that of Dollard and Miller. He does not feel that the neurotic's difficulties arise because of overrestraint and unrealistically high moral standards, and he does not believe that superego functions or values can be influenced by psychotherapy. The study reported here bears on this issue. In the main, we undertook to test the following hypotheses: that values do change in psychotherapy, particularly moral values since these are thought to be at the center of neurotic conflicts; that the patient learns to accept the moral values of the therapist, and that alleviation of psychological distress accompanies this new learning. It was hypothesized too that values such as those found in the Allport-Vernon Scale of Values would not be systematically influenced by psychotherapy, since these would not ordinarily be at issue either in the neurosis or its treatment.

Procedure

Patients at the Henry Phipps Psychiatric Clinic were accepted for the study only when they and their doctors agreed to take all the necessary tests. No patients diagnosed as psychotic were included. Some high school education was also made a requirement for inclusion since many of the items on the tests could not be understood by people of low education or intelligence.

Many patients who were considered eligible never completed the experiment. Some were too upset to take the tests or had had shock therapy. Others took the tests initially but left against advice shortly thereafter, either before arrangements could be made for retesting, or else with a flat refusal to be retested. As it turned out, it took two years to accumulate twelve patients who were tested before and after therapy. Of these, nine were hospitalized or "house patients," while the remaining three were treated on an outpatient basis. Six were male, six female. Diagnoses included psychoneurotic, personality, psychophysiologic, and adjustment disorders. Ages ranged from 18 to 46, with a mean of 29.5 years. Length of treatment ranged from 3 weeks to one year, with a mean of 5 months.

At some point early in therapy, after fewer than six therapeutic interviews, the patients were given four tests: Frank's Symptom-Disability Check List; the Allport-Vernon-Lindzey Scale of Values (1); the Butler-Haigh (2) Self Concept items, from which was derived a modification of the Dymond Adjustment Scale (5); and a Moral Values Q sample devised by the author. These tests were repeated at the end of treatment. The therapists, all psychiatric residents at the Henry Phipps Psychiatric Clinic, were given the Allport-Vernon-Lindzey Scale of Values and the Moral Values Q sample. These tests were not readministered to the therapists at the end of treatment since it was not felt that they would show any predictable change. It could have been hypothesized that the therapist also assimilated some values of the patient, but when it is realized that a therapist treats many patients at the same time, each holding a different set of values, then it seems doubtful that he can assimilate the values of all in any readily measurable way. At any rate, it was assumed in this study that the therapist's values were stable.

At the end of treatment, each patient was interviewed by the author who had no knowledge of how the final and initial test results compared. The interview was open-ended, the patient being given an opportunity to air all thoughts and feelings bearing on his therapeutic experience. Almost all welcomed the opportunity to tell how they felt. During the interview, the author directed the discussion to the following questions: How did the patient feel as compared to when he began therapy? If he felt better, how was he better? If worse, how worse? If he felt better, what did he think helped him to feel better? How did he feel about his therapist? What did he feel about therapy, terminating therapy, the future? No mechanical recording was made of the interview since it might have reduced rapport and the willingness to speak freely. Immediately after the interview, a detailed report of it was made by the author. Bias could not have entered into the reporting without prior knowledge of the test results. At the end of the experiment, all twelve interviews were rated independently by three judges on a 7-point scale as to whether the patients benefited or not by their therapeutic experience.

The Allport-Vernon-Lindzey Scale of Values was administered in the usual way. The Symptom-Disability Check List was developed by Dr. Jerome D. Frank and his colleagues in the course of research on the response of patients to individual and group therapy. It has not yet been published. It consists of 41 symptoms such as "headaches," "feeling blue," "unusual fears," etc., which are rated by patients on a 4-point scale of distress, and 12 additional items rated on a 6-point scale of severity.

The Butler-Haigh set of 100 statements, intended by its authors as a Q sample, was administered to our patients by asking them to sort the statements into two piles: 1. Do *not* describe you well; are *not* so true of you; are more *unlike* you than like you; do *not* apply to you so well. 2. Describe you fairly well; are true of you for the most part; are like you more than unlike you; apply to you more than not. Since we wished by this procedure to obtain a score based on the Dymond Adjustment Scale, this dichotomous sorting was quite adequate and it was a simple task for the patients.

The Moral Values Q sample consists of 60 statements based on three main areas of behavior around which psychological conflicts commonly arise: sex, aggression, and authority. Each area in turn is subdivided so that two generally contrasting sets of attitudes are represented with respect to each area. Thus, there are six categories:

1. *Sex-rigid:* sex is seen as evil, perhaps a necessary evil, dirty, to be shunned, postponed, or suppressed.
2. *Sex-free:* sex is seen as good, healthy, natural, to be accepted, enjoyed, understood.
3. *Antiaggressive:* acts of aggression, violence, or pain infliction are thought of as wicked, to be avoided, suppressed, suffered.
4. *Aggressive:* aggression is thought of as necessary, rewarding, to be accepted, expressed.
5. *Disciplinarian:* authority, discipline, order, and regulation are seen as absolute, necessary, desirable, to be accepted, supported.

6. *Libertarian:* authority is seen as a relative matter, with individual responsibility and reason primary.

Moral values were thought to involve behavior which is seen as right or wrong, good or bad, permissible or not permissible, worthy of reward or punishment. Items were sought which would reflect such attitudes. Ten items selected from therapy protocols and from other sources were written for each category. Controversial issues which aroused strong feelings were given priority since it was felt that conflicts would more generally arise around such issues. The statements are as follows:

Sex-Rigid

3. It is a wife's duty to have intercourse with her husband.
10. Girls should come to the marriage bed as virgins.
15. It is bad taste to tell dirty jokes in mixed company.
22. Men who have sex relations outside of marriage are low and contemptible.
27. Divorce is never justified.
33. People should not marry before they are twenty-one.
39. Sexual intercourse should be had only for the purpose of having children.
46. The best and truest love is one without sexual thoughts and feelings.
52. Pulp magazines featuring sexy love stories should be abolished.
57. Masturbation is a sin.

Sex-Free

4. It's perfectly all right for teen-agers to kiss and neck.
9. Men ought to have some sexual experience before marriage.
16. Sexual intercourse provides one of life's most beautiful experiences.
21. A woman may be justified in having another man if her husband is not satisfying her sexual needs.
28. Prostitution should be legal.
34. Unnatural sex acts should not be considered a crime.
40. Sex-play before intercourse makes it more satisfactory.
45. Sex education should be taught in public schools.
51. One should expect one's wife or husband to be attracted to handsome members of the opposite sex.
58. If young children play with their sex organs, parents need not be worried or concerned.

Antiaggressive

1. If someone strikes you on one cheek, turn the other.
12. Using live animals for medical research is wrong.
13. Judge not lest ye be judged.
24. One should always try one's best to get along with others.
25. It is wrong to hate.
31. Capital punishment is wrong.
37. If you're angry, count ten and try to let the feeling go away.
48. Wars are never justified.
54. The meek shall inherit the earth.
55. Violence breeds violence.

Aggressive

2. A man who lives by the sword should die by the sword.
11. Might makes right.
14. One should say what one thinks about other people.
23. If people annoy you, either let them know it or don't bother with them.
26. Some people are no damn good.
32. Let the punishment fit the crime: an eye for an eye.
38. To the victor belong the spoils.
47. A nation that will not defend its rights or its honor deserves not to survive.
53. The race is to the swift and the strong.
56. You have to use brute force to set some matters straight.

Disciplinarian

5. Labor unions should be strictly regulated.
8. One should always be busy and keep one's mind occupied.
17. Children should do as they are told.
20. Parents act in the best interest of their children.
29. Parents sacrifice a lot for their children.
35. "Thou shalt honor thy father and mother" is a commandment which should always be observed.
41. Everyone should belong to some church.
44. What our country needs most are great leaders.
50. A soldier should always obey orders first, ask questions afterward.
59. Our future lies in the hands of fate.

Libertarian

6. There may be times when one should break the law.
7. People should take time out to relax and think about themselves.
18. Parents should always tell children why things should or should not be done.
19. Many parents do not really love their children.
30. The trials of parenthood are more than equalled by its rewards.
36. Parents often deserve to lose the respect of their children.
42. God is created by man in his own image.
43. Famous leaders owe their greatness to the ones they lead.
49. An employee should express his opinion, even when the boss disagrees.
60. A person must take responsibility for his own life.

Table 1

Rho Correlations of Criteria of Improvement

Criteria	*Rho*	*df* (*N*−2)	Significance levels†
Interviews and SDCL	.61	10	.038
Interviews and Adjustment Score	.85	9*	.001
SDCL and Adjustment Score	.48	9*	.144

* Patient D, the first in our series, had his initial testing before we had decided to use the Dymond Adjustment Scale, and so no measure of change on this criterion was available for him.

† Significance levels were determined according to procedures suggested by Kendall (8).

Patients and therapists were required to sort the 60 statements into eleven piles based on the strength of their belief or disbelief regarding each statement. The number of statements to be placed in each pile was as follows: 1, 2, 4, 7, 10, 12, 10, 7, 4, 2, 1. The rationale underlying these procedures has been discussed at length by Stephenson (15).

Results

The main hypothesis of the study was that patients who improved would tend to modify their system of moral values in the direction of their therapists' moral values. A review of the literature makes it apparent that the problem of criteria of improvement is an unsolved one. We used three possible criteria: change on Frank's Symptom-Disability Check List (SDCL); change on Dymond's Adjustment Scale, based on the patient's sortings of the Butler-Haigh statements; and judges' ratings based on the post-therapy interviews. The latter are admittedly crude, but they have the advantage that in the interview the patient could speak freely, and that where he was motivated to create an impression of improvement or lack of it, this could be reflected in what he said and taken into account by the judges.

Patients were ranked on all three criteria, the rankings intercorrelated to determine the comparability and stability of the measures, as shown in Table 1.

From Table 1 it can be seen that the judges' ratings based on the interviews correlated significantly with ratings based on both the Symptom-Disability Check List and the Adjustment Scale, but that the latter did not correlate significantly with each other. It thus appeared that the judges' ratings were the best single index of improvement.

To determine the direction of change in patients' moral values, each patient's initial and

Table 2

Judges' Improvement Ratings and Moral Values Correlations

Patient	Mean of judges' ratings*	Rank of judges' ratings	Correlation of patients' initial and final sortings	Correlation of patients' sortings with therapists' sortings: Initial 1	Final 2	Z_2-Z_1	Rank Z_2-Z_1
A	5.3	1.5	.47	.108	.371	.281	1
B	5.3	1.5	.72	−.003	.200	.206	3
C	4.8	3	.86	.664	.653	−.019	7
D	4.7	4.5	.48	.443	.535	.122	5
E	4.7	4.5	.84	.398	.402	.005	6
F	4.2	6	.70	.173	.399	.248	2
						Sum .843	
G	4.0	7	.67	.192	.015	−.179	11
H	3.0	8	.58	.345	.226	−.130	9
I	2.8	9	.54	.451	.406	−.055	8
J	2.3	10	.51	.245	.398	.172	4
K	2.2	11	.59	.268	.131	−.143	10
L	1.7	12	.62	.200	−.011	−.214	12
						Sum −.549	

* Ratings of Improvement: much improved = 6; moderately improved = 5; slightly improved = 4; unimproved = 3; slightly worse = 2; moderately worse = 1; much worse = 0.

final sortings of the moral value items were correlated with the therapist's sorting. Table 2 shows the main data bearing on the hypothesis.

The rho correlation of judges' ratings of improvement and change in direction of therapist's moral values equals .68, which, for $df = 10$, has a significance level of .016. If we compare the changes in moral values of those above the median improvement rating with those below the median, we find $t = 3.013$ which is also significant at the 2 per cent level. Thus, the main hypothesis appears to be supported by the data.

However, the changes in the moral values, when considered in relation to the therapist's values, are not very large. This could mean that although the system of values is modified, it is not profoundly altered. An alternative explanation is that greater changes could not be expected since only two of the patients were rated as more than moderately improved. In one of these, Patient A, the changes in moral values were quite large.

It appears too that the hypothesis needs to be expanded to account for the fact that patients who were unimproved or worse tended to move *away* from the therapist's value system, so that changes in moral values may be effected in both directions by the therapeutic situation.

The occurrence of exceptions to the hypothesis, Patients C, E, G, and J, indicated that other factors may be involved as well as changes in moral values to account for improvement or lack of it. Patient C thought that what helped him as much as anything was the discovery that he was smarter than his therapist. According to patient E, who had been experimenting with paranoid ideas and wanted to "sign out" of the hospital, he was told by his therapist that if he signed out he would have to be committed to a state hospital since as long as he believed his paranoid ideas he was insane. This was a "jolt" to him. Afterward, he began to question his "beliefs," found out what purposes they served him, and finally gave them up. Thus, therapy in these cases took an unusual course where conflicts around moral values were not at issue as far as the patients were concerned. Patient G "signed out" against advice. Her husband, in service, was being transferred to another part of the country and she wanted to go with him, fearing to be left behind. She was thus bent on convincing everyone of her "improvement" so that they would let her go. When asked what she thought had helped her, she said that it was being around the other patients, especially those on the first floor (psychotics) and learning from them how she should be or should not be. She was discharged as unimproved by her therapist, rated as slightly improved by all three judges, and made herself out to be considerably improved on the Symptom-Disability Check List and the Adjustment Scale. Patient J spontaneously remarked that therapy had taken away her former values and had not replaced them with anything. She felt "confused" at the time of her discharge but was planning to continue therapy on an outpatient basis. Thus, the effect found here might be one of therapy-in-transition.

To test whether similar changes would occur with respect to values of the type measured by the Allport-Vernon-Lindzey Scale of Values, the score obtained by the patient in each of the six scale categories, both before and after therapy, was compared with his therapist's scores in the same categories by means of the *D* technique (3). Four patients' profiles on this scale became more similar to their therapists', eight less similar. Change in similarity to the doctor's profile showed a rho correlation of .18 with amount of improvement. Since this is not significantly greater than zero, we cannot assume any support for the hypothesis of a relationship between changes in such values and improvement. This accords with our expectations, since it was believed that such values were not likely to be importantly involved either in neurosis or therapy.

Summary and Conclusion

Twelve patients at the Henry Phipps Psychiatric Clinic were given a battery of four tests before and after psychotherapy. Their therapists were given two of these tests. Patients who improved tended to revise certain of their moral values in the direction of their therapists', while the moral values of patients who were unimproved tended to become less like their therapists'. This was not found in the case of values such as those of the Allport-Vernon-Lindzey scale.

It was thought that the latter type of values showed no systematic change with respect to improvement since it was not likely to be at issue in the neurosis itself or in its treatment. Changes in moral values centering around sex, aggression, and authority were thought to occur because these issues are commonly involved in patients' conflicts.

If patients who improve change these moral values to accord more with their therapists' moral values, it seems likely that the patients have learned to accept part of the latter's value system as their own. Most therapists, however, seem to take great precautions to avoid influencing their patients' values in any way at all. Shoben (14) emphasizes that the patient must evolve his own set of moral values and that the therapist must not engraft his own values upon him. It may be that the therapist communicates his values to the patient in many unintended, subtle ways, even when trying to avoid so doing. The patient, who is often sensitized to the therapist's every word and inflection, may be able to receive these communications, and because of his trust, admiration, and respect for the therapist, may permit himself to be influenced by them.

It seems likely, too, that there is a direct connection between one's self concept and one's system of moral values. These values are an important part of the standards by which the self is judged, yet not only the self but others as well. Thus, Sheerer's (12) finding that there is a direct relationship between change in acceptance of self and change in acceptance of others is predictable when it is considered that both self and others are being judged in the light of a changed set of values.

The findings would not seem to support Mowrer's notion that "superego functions or values can be influenced very little, if at all, by psychotherapy." He places emphasis instead upon "reestablishing normal *communication* between ego and superego and upon changing *ego functions* rather than superego values or standards" (9, p. 92). It may be that the changes found in this study are primarily evidences of better communication between ego and superego, but perhaps it is necessary for us to be able to differentiate operationally these various structural concepts before we can render any verdict about processes affecting them.

Received April 5, 1955.

References

1. Allport, G. W., Vernon, P. E., & Lindzey, G. *Study of values*. Boston: Houghton Mifflin, 1951.
2. Butler, J. H., & Haigh, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 55–75.
3. Cronbach, L. J., & Gleser, Goldine C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
4. Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
5. Dymond, Rosalind F. Adjustment changes over therapy and self-sorts. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 76–84.
6. Eysenck, H. J. The effects of psychotherapy: an evaluation. *J. consult. Psychol.*, 1952, **16**, 319–325.
7. Fiedler, F. E. Quantitative studies on the role of the therapists' feelings toward their patients. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald Press, 1953. Pp. 296–315.
8. Kendall, M. G. *Rank correlation methods*. London: Griffin, 1948.
9. Mowrer, O. H. Neurosis and psychotherapy as interpersonal processes: a synopsis. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald Press, 1953. Pp. 69–94.
10. Mowrer, O. H. Neurosis, psychotherapy, and two-factor learning theory. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald Press, 1953. Pp. 140–149.
11. Rogers, C. R. Some directions and end points in therapy. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald Press, 1953. Pp. 44–68.
12. Sheerer, Elizabeth T. The relationship between acceptance of self and acceptance of others. *J. consult. Psychol.*, 1949, **13**, 169–175.
13. Shoben, E. J., Jr. Some observations on psychotherapy and the learning process. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald Press, 1953. Pp. 120–129.
14. Shoben, E. J., Jr. Anxiety vs. immaturity in neurosis and its treatment. *Amer. J. Orthopsychiat.*, 1955, **25**, 71–80.
15. Stephenson, W. *The study of behavior*. Chicago: Univer. of Chicago Press, 1953.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Projection, Self Evaluation, and Clinical Evaluation of Aggression[1]

Anthony Davids,[2] Andrew F. Henry,[3] Charles C. McArthur,
and Leo F. McNamara[4]

Harvard University

What relation do psychologists expect between the way a need is expressed in projective tests and the way it is expressed in behavior? According to an hypothesis of "direct projection," the two expressions of the same need should be parallel; while according to an "hydraulic theory," the amounts of expression in fantasy and behavior should be inverse. One might also make different predictions based upon more complex relations that have been proposed to exist between fantasy expression and behavioral expression of needs. Whatever theoretical position one chooses, empirical examination of hypotheses is facilitated by assessment of some need that frequently gains expression both in fantasy and behavior. One of the most common, as well as one of the most important, both theoretically and practically, is "n Aggression."

Within n Aggression, it is customary to differentiate between extrapunitive aggression and intropunitive aggression. In his original work on the Thematic Apperception Test (TAT), Murray (6) pointed to the clinical importance of the subcategory of aggression termed "n Intraggression." Rosenzweig, also, has demonstrated the theoretical and empirical utility of this dichotomy between inward- and outward-directed aggression (7, 8). Moreover, recent research (5) has shown cardiovascular reactions to be related to "Anger In" and to "Anger Out" emotional reactions in response to stress. And these concepts have proved useful in sociological studies of suicide and homicide (4).

The present experiment represents an attempt to further explore relations among various measures of n Aggression. Specifically, we will relate measures of aggression derived from self ratings, clinical ratings, and projective protocols.

Method

Subjects. The subjects (*S*s) were 20 male college students who volunteered to serve as paid *S*s in an intensive investigation of personality. They participated in approximately 40 hours of psychological examination, consisting of psychodiagnostics, experiments, and interviews, conducted by a team of investigators.

Projective ratings. The TAT, administered individually to each *S*, was electrically recorded and transcribed verbatim. An experienced clinical psychologist, who had not administered the test and who was not personally familiar with the *S*s, analyzed the stories given in response to cards 3, 6, 7, 8, and 13 of the male series. On the basis of his analysis, he rated the *S*s in regard to both *amount* and *direction* of aggression. That is, upon analyzing each *S*'s TAT stories, the psychologist rated him as either high or low on aggression as well as rating him as either characterized by anger directed inward or outward.

Self evaluation. Included in a comprehensive inventory of self ratings was a scale de-

---

[1] This study is an outgrowth of a program of research conducted at the Harvard Psychological Clinic under the direction of Professor Henry A. Murray and supported by a grant from the Rockefeller Foundation. The authors are indebted to the following institutions for facilitating this report: the National Institute of Mental Health (Grant M-700), Public Health Service, the Laboratory of Social Relations, and the Study of Adult Development at Harvard University.

[2] Now at Brown University and Emma Pendleton Bradley Home.

[3] Now at Vanderbilt University.

[4] Now at the Queen's University of Belfast.

Table 1

TAT Rating and Self Rating on Amount of Aggression

Self rating	TAT	
	High	Low
High	6	5
Low	6	3

Fisher's $p = .47$.

Table 3

TAT Rating on Direction of Aggression and Self Rating on Amount of Aggression

Self rating	TAT	
	Out	In
High	10	1
Low	1	8

Fisher's $p = .001$.

signed to measure conscious avowal of aggression. The *S* was provided with a 6-point rating scale running from "much less than most people" to "much more than most people," and instructed to indicate his general level of aggression. For purposes of the present investigation, *S*s who rated themselves as more aggressive than most people were classified as high on aggression and those who rated themselves as less aggressive than most people were classified as low on aggression.

Clinical evaluation. An experienced clinical psychologist, who had contact with the *S*s over approximately an 18-month period, rated each *S* in regard to both *amount* and *direction* of aggression. That is, utilizing information gleaned from his personal contacts and from his impressions derived from the *S*s' responses to a battery of tests and experiments, the psychologist rated each *S* as either high or low on aggression and classified him as either extrapunitive or intropunitive. It should be emphasized that the TAT was not in the battery of tests analyzed by this psychologist and that he was not the same person who made the projective ratings.

Statistical treatment of data. Association between the various sets of ratings is presented in the form of 2 × 2 contingency tables. Because of the low expected frequencies in the cells of these tables, Fisher's (3) exact method for determining significance of association is employed rather than the more conventional chi-square technique. Thus, the statistic utilized to report significance of findings is Fisher's p.

Results

From the results presented in Table 1, it appears that the relation between TAT ratings and self ratings on amount of aggression is essentially random. This finding might be regarded as negative evidence for the validity of the TAT. Additional evidence suggesting lack of TAT validity might be drawn from the fact that the relation between TAT ratings and clinical ratings on amount of aggression, shown in Table 2, is also random. Upon further analysis, however, it appears that the seeming lack of validity of the TAT may be due to an artifact.

The reason is suggested by the findings shown in Table 3. Here it becomes evident that there is an almost perfect association, significant beyond the .001 level, between the *amount* of aggression avowed in self ratings and the *direction* of aggression as assessed in thematic apperceptions. Individuals with pronounced n Intraggression, according to an experienced psychologist's interpretation of the TAT's, rated themselves as low on n Ag-

Table 2

TAT Rating and Clinical Rating on Amount of Aggression

Clinical rating	TAT	
	High	Low
High	5	4
Low	7	4

Fisher's $p = .54$.

Table 4

Clinical Rating on Direction of Aggression and Amount of Aggression

Clinical rating	Clinical rating	
	Out	In
High	8	1
Low	3	8

Fisher's $p = .01$.

Table 5

TAT Rating on Direction of Aggression and Amount of Aggression

	TAT	
TAT	Out	In
High	5	7
Low	5	3

Fisher's $p = .33$.

gression. This finding suggests that the men evaluated themselves under the assumption that only outward-directed aggression was to be considered in estimating the amount of one's own aggressiveness.

In this regard, it is noteworthy, as evidenced by Table 4, that the clinician also tended ($p = .01$) to equate a high rating on amount of aggression with a rating of "Anger Out." These statistically significant findings are reinforced by the observation that out of the entire group of *S*s only one individual was rated high on aggression by both himself and the clinician and was also rated clinically as directing his aggression inward. The clinician was fully aware that need aggression can gain expression in two directions. Yet a major difference between the clinical ratings and the TAT ratings was that for the clinician high n Aggression was associated with "Anger Out" while for the TAT, as shown in Table 5, high and low n Aggression occurred independently of the rating on "Anger Out" or "Anger In." As a result, of course, the projective ratings bear a random relation to the clinical ratings.

Discussion

On the basis of these findings, one might conclude that the TAT measure of inward-directed aggression is invalid. The alternative conclusion, however, is that the TAT is making a unique contribution to the case record. It seems to us that the latter interpretation is more plausible. In American culture there are many legitimate means for expressing "Anger Out," even in quite "proper" situations. For example, TAT stories of murder and assault reflect, in an extreme degree, a need that in everyday life may find less drastic expression in such things as changing an appointment, being harshly critical of the work of others, or engaging in heated discussion of politics. Often the TAT measure of n Extraggression will do no more than confirm the presence and extent of influence of this need that is displayed in many facets of the case record. In fact, the person himself may be well aware of his aggressive feelings toward people and situations in his environment, and, upon direct questioning, he may be both able and willing to present an accurate account of this aspect of his personality make-up.

By contrast, however, n Intraggression does not have many outlets for expression in everyday behavior. In fantasy, one's heroes may be victims of fate or may commit suicide, but in ordinary life one has little recourse beyond an occasional "accident" or a self-critical remark. Moreover, mild degrees of intraggressive behavior are not defined as such by society at large. Therefore, it seems likely that the intraggressive need will *not* be written bold over all the case record. Here is where an "hydraulic" function is served by the TAT—it presents an opportunity to express in fantasy a need that rarely gains behavioral expression.

It seems that we are left with a slight variation of an old rule. This version is, "For needs with behavioral channels to expression, the TAT and behavior are parallel; while for needs with few behavioral channels, measures derived from the TAT and from behavior are inverse." However, this formulation requires further consideration. Actually, both aspects of this formulation can be viewed as special cases of the general rule that "TAT projections are directly parallel to other behavior." The narrator whose TAT heroes feel n Intraggression possesses such a need himself in real life and it influences his actions, but society prescribes that he can act out this particular need with less visible behavior. One result of this is that many men who have in their TAT protocols the theme "Aggression expressed outward followed by aggression directed inward" were rated predominantly as "Anger Out" and "High" in n Aggression both by the psychologist and by themselves. It seems that only the first part of the theme was readily translated into visible behavior.

Occasionally, however, a shrewd observer managed to deduce the complete theme on the basis of subtle behavioral signs. For example, one person, whose TAT heroes all evidenced a strong need to direct their aggression outward but were subsequently "reluctant" to express this feeling, was rated correctly by an observer who said that the subject was "so suspiciously nice I had the feeling there was something going on there."

Thus it seems plausible that the relation between projected needs and behavioral needs is direct parallelism. Yet a projective test may add something unique to a case record because certain needs find more condoned expression in samples of fantasy than in samples of any other kind of behavior. When it is equally easy to act on a given need both directly and in projective situations, however, it seems likely that knowledge of the need will not be greatly increased by adding a projective test to one's battery of assessment techniques. For such needs one may, at least when examining healthy people who have no motivation for distorting the data, do quite as well by eliciting an *S*'s direct report as by utilizing a complex projective method of personality assessment. This point has recently been stated rather convincingly by Allport (1) and has received empirical confirmation in an experiment reported by Davids (2). This notion is further substantiated by the present finding shown in Table 6, of significant association ($p = .01$) between the *S*s' self evaluations of whether they were high or low on aggression and the clinician's evaluation, based on many hours of assessment, of their standing on this attribute.

Although he believes that direct methods are as effective as projective methods when applied to normal *S*s, Allport (1) states that the TAT and other projective techniques often make an important contribution in the assessment of neurotics who are apt to curtail their self-expressions defensively. We would extend his statement and say that, since the culture curtails the expression of *certain* needs in almost everyone, the TAT will contribute to increased understanding of these needs in normals as well as in neurotics. In fact, since the behavioral expression of such needs may be even more inhibited in emotionally healthy people, who are more likely to be "culturally normal," the relative contribution made by projective methods may be even greater when these needs are assessed in normals than when they are assessed in neurotics.

Table 6

Self Rating and Clinical Rating on Amount of Aggression

Clinical rating	Self rating	
	High	Low
High	8	1
Low	3	8

Fisher's $p = .01$.

Conclusions

In the present study, both the self and the clinician, we have inferred, were measuring and evaluating n Extraggression—a need that has many culturally sanctioned channels of expression. Under these conditions the projective test behavior and the rest of the recorded behavior are concordant, and, in this sense, nothing unique is added by the projective method. However, it seems to us, the unique advantage and contribution of projective techniques becomes evident when one examines n Intraggression—a need whose expression in behavior other than fantasy is culturally made difficult.

Received March 18, 1955.

References

1. Allport, G. W. The trend in motivational theory. *Amer. J. Orthopsychiat.*, 1953, 23, 107–119.
2. Davids, A. Comparison of three methods of personality assessment: direct, indirect, and projective. *J. Pers.*, 1955, 23, 423–440.
3. Fisher, R. A. *Statistical methods for research workers.* London: Oliver & Boyd, 1936.
4. Henry, A. F., & Short, J. F. *Suicide and homicide.* Glencoe, Ill.: The Free Press, 1954.
5. King, S. H., & Henry, A. F. Aggression and cardiovascular reactions related to parental control over behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 206–210.
6. Murray, H. A. *Thematic Apperception Test Manual.* Cambridge: Harvard Univer. Press, 1943.
7. Rosenzweig, S. The experimental measurement of types of reaction to frustration. In H. A. Murray (Ed.), *Explorations in personality.* New York: Oxford Univer. Press, 1938. Pp. 585–599.
8. Rosenzweig, S., Fleming, Edith E., & Clarke, Helen J. Revised scoring manual for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1947, **24**, 165–208.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Repression and the "Return of the Repressed"[1]

John H. Flavell
Clark University[2]

Zeller (4, 5) recently conducted an experiment unique among laboratory studies of repression. Not content merely to demonstrate again that something akin to repression of threat-associated material may occur in a memory experiment, Zeller further showed that a reminiscence or reinstatement of this forgotten material can be seen when the threat is later disassociated from it. The threat used in his study was that of an induced feeling of failure in the performance of a motor task. If Zeller has in fact created an experimental analogue of Freud's "return of the repressed" (1), his study deserves to be considered another "first" among experimental investigations of psychoanalytic phenomena.

It is the purpose of the present study to attempt verification of Zeller's findings. The present design incorporates two additional refinements, however. First, in order to approximate more closely than Zeller has done the clinical conditions under which repression and return of the repressed would ordinarily be expected to occur, a threat to personal adjustment was substituted for Zeller's threat of failure on a motor task. Second, an attempt was made to discover whether repression and return of the repressed are selective or general in their effects, i.e., whether or not the repressive influence of threat is mainly confined to that memory material most closely associated with the threat. In view of Korner's (2) experimental findings, the expectation would be that threat to personal adjustment would be selective rather than general in its effects. Therefore, the study proposes the following three hypotheses: (*a*) a threat to personal adjustment will produce a significant decrement in the recall of verbal material associated with this threat; (*b*) a partial reminiscence of this forgotten material will be in evidence after threat conditions are removed; (*c*) forgetting and reminiscence induced by threat and threat-removal will tend to be specific to that verbal material most closely associated with threat.

Method

Subjects

Thirty-eight college freshmen and sophomores served as volunteer *S*s. Prior to the experiment proper these *S*s were divided into two equal-sized groups, roughly equated for nonsense syllable recall. The *S*s were unselected for sex, age, or intelligence.

Procedure

All *S*s were seen individually for four consecutive daily sessions.

Session I. The *S*s of both groups were shown a series of twelve nonsense syllables, exposed for two seconds each, under instructions to memorize for future recall. Two more exposures of the twelve syllables followed. The *S*s were then instructed to write a one-word association to each syllable as it was presented a fourth time. Recall Test 1 followed. In this and the four subsequent Recall Tests the *S*s were given four minutes in which to write down as many of the twelve syllables as they could remember.

Session II. For the experimental group *S*s, Session II proceeded as follows. When the *S* entered the room he was introduced to a psychology professor[3] and told that the latter had kindly consented to "interpret" the *S*'s one-word associations of the previous day as "normal" or "abnormal" with respect to "those of the general college student population." The *E* then read aloud the twelve nonsense syllables and the associations the *S* had previously given to each. The professor commented "abnormal" or "normal" after each syllable-association pair in a prearranged sequence which was identical for all *S*s. In this manner it was insured that six of the nonsense syllables were always indirectly associated with a comment of "normal" and the remaining six with "abnormal." The procedure was then repeated "to make sure the interpretations were correctly recorded." Following this, the professor left the room and Recall Test 2 ensued, in which the *S* again wrote down all the nonsense syllables he could remember. The control group *S*s were likewise given two verbal presentations of the syllable-association pairs followed by Recall Test 2. However, the professor was not present and no interpretations of associations were made. Rather, the two presentations were given under the pretext first, of making sure that the associations were correctly recorded and, in the case of the second presentation, to ascertain whether the *S*s were satisfied with these particular associations. Thus, for the experimental group, Session II consisted of two "incidental learning" trials, personality threat indirectly associated with six of the twelve nonsense syllables, and Recall Test 2. For the control group, the procedure was essentially identical except that no threat was introduced.

---

[1] This study formed the basis for a Master's thesis submitted to Clark University in 1952. It was carried out under the direction of Dr. Thelma G. Alper, for whose invaluable assistance the author wishes to express his gratitude.

[2] Now at the University of Rochester.

[3] Dr. Thelma G. Alper.

Session III. For the experimental *S*s, Session III commenced with Recall Test 3. The psychology professor then entered and explained to the *S* in some detail that her interpretations of the day before were completely bogus, that the whole affair had been staged for experimental purposes to test the effect of stress on performance, etc. In order both to assess the effectiveness of the stress and to provide some catharsis for the *S*s, each *S* was asked to comment on his reaction to the interpretations. About half of the *S*s frankly stated that the interpretations had been very disturbing to them. The professed lack of anxiety on the part of the remaining *S*s was in most cases belied by the presence of observable tenseness, facial grimaces, etc., and by the obvious relief most *S*s showed when told they had been victims of an experimental hoax. Recall Test 4 immediately followed this "threat removal" interview. The control *S*s were likewise given Recall Tests 3 and 4, separated, however, by an empty interval of approximately the same duration as the experimental *S*s' interviews.

Session IV. The last session, identical for both groups, consisted solely of Recall Test 5. It is thus evident that, for the experimental *S*s, the design of the study was essentially as follows: induction of threat—immediate recall—24-hour recall; and, removal of threat—immediate recall—24-hour recall. The control group followed the same procedure except that personality threat was nowhere introduced.

Results and Discussion

Figure 1 shows "normal," "abnormal," and "total" nonsense syllable retention curves for the two groups. The following within-group changes in mean recall scores yield statistically significant *t*'s ($p < .05$) for a one-tailed test: (*a*) C group increase in "total" recall from Recall Test 1 to Recall Test 2; (*b*) E group decrease in "total" recall from Recall Test 1 to Recall Test 2; (*c*) E group increase in "total" recall from Recall Test 3 to Recall Test 4; (*d*) C group increase in "normal" recall from Recall Test 1 to Recall Test 2. The fact that the E group mean "total" recall dropped significantly on Recall Test 2 whereas the corresponding C group score actually showed a significant increase strongly suggests that the hypothesized threat-induced repression did occur. Likewise, the significant increase in E group "total" recall in Recall Test 4 confirms the hypothesis of a remission of repressive effects following removal of threat. It was therefore concluded that our experimental findings concur with those of Zeller, both with respect to repression proper and with respect to what has been termed "a return of the repressed."

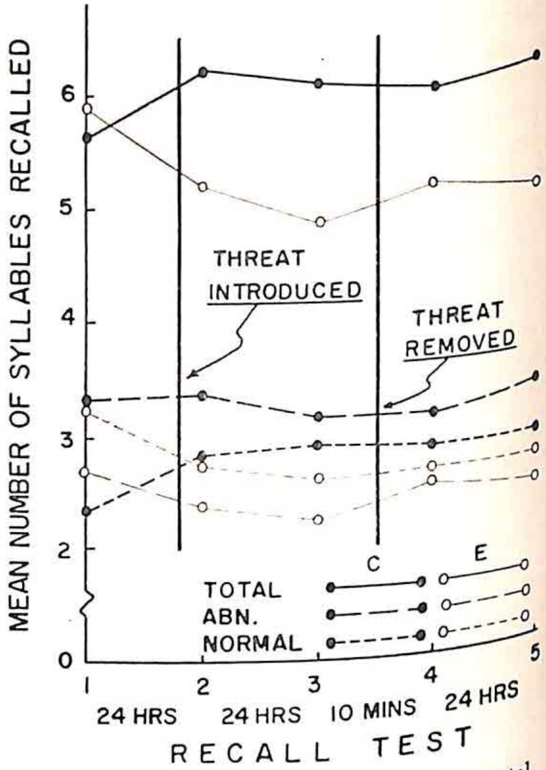

Fig. 1. Nonsense syllable recall for experimental (threatened) and control (nonthreatened) groups.

The hypothesis that the repression process would be selective rather than general in its effects does not appear to be confirmed by the data. It is apparent from Figure 1 that nonsignificant decrements in both "normal" and "abnormal" syllable recall contribute almost equally to the significant "total" E group decrease from Recall Test 1 to Recall Test 2. Similarly, the increase in E group "abnormal" recall from Recall Test 3 to Recall Test 4, although perhaps impressive looking, is not statistically significant. It is possible that repression tends to be selective or general depending upon the extent to which threat-associated material is clearly articulated from material not associated with threat. In Korner's study, the threat was unequivocally associated with certain discrete and self-contained stories and just as unequivocally not associated with other such stories. In the present study threat was indirectly associated with certain nonsense syllables imbedded in a series of similar syllables and thus differentiation may have been difficult. Hence, the spread of repressive effects to material not specifically associated with threat in the present study may be an example of what Scheerer (3) has referred to as "generalization by default."

Summary

The present study investigated the effects of introduction and subsequent removal of threat to personal adjustment on the recall of nonsense syllables, some of which were indirectly associated with the threat. The experimental results confirmed the hypotheses that threat would produce a decrement in recall analogous to defensive repression and that subsequent threat removal would cause a partial reminiscence of forgotten material akin to Freud's "return of the repressed." The hypothesis that threat would be selective rather than general in its effect upon recall was not confirmed.

Received April 18, 1955.

References

1. Freud, S. Repression. In *Collected papers.* Vol. IV. London: Hogarth, 1925. Pp. 84–97.
2. Korner, I. N. Experimental investigation of some aspects of the problem of repression: repressive forgetting. *Teach. Coll. Contr. Educ.*, 1950, No. 970.
3. Scheerer, M. Cognitive theory. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge: Addison-Wesley, 1954. Pp. 91–142.
4. Zeller, A. F. An experimental analogue of repression. I. Historical summary. *Psychol. Bull.*, 1950, **47**, 39–51.
5. Zeller, A. F. An experimental analogue of repression. II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, **40**, 411–422.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Further Validation of the Ego-Strength Scale[1]

Robert D. Wirt[2]
Veterans Administration Hospital, Minneapolis

The Minnesota Multiphasic Personality Inventory was administered to a population of hospitalized patients who had been treated by psychotherapy alone. These were scored for Barron's Ego-Strength scale (1). The results confirmed, for a hospitalized sample, Barron's findings for clinic samples: the Ego-Strength scale has good predictive power for response to psychotherapy.

All patients admitted to the Minneapolis Veterans Administration Hospital during the period February 1, 1953, to February 1, 1954, routinely were administered the MMPI. Of these 535 cases, the 225 in whom there were no organic complications and who did not receive physiological adjuncts to treatment were selected for study. Upon discharge and without knowledge of the *Es* score, each patient was rated by his psychiatrist as Unimproved, Improved, or Greatly Improved. The chief of the neuropsychiatric service, who acted as supervisor for all cases treated by psychotherapy alone, also rated the patients using these categories. The 203 cases who received identical ratings and who were treated by psychotherapy alone were selected for the sample. While only one item (No. 45) significantly separated the Greatly Improved from the Unimproved patients, and four items (Nos. 95, 109, 355, 554) tended in the direction opposite from that found by Barron, the scale as a whole separated the groups at better than the .05 level of probability. The mean raw score for the sample was 42.3 and the standard deviation was 12.46. It was possible to construct a conversion table of *T* scores for use with hospitalized male patients.

No patient in the Unimproved Group exceeded the median of the patients in the Greatly Improved Group. If a raw score *Es* value of 50 is used as the cutting point, there are no false positive predictions, i.e., all patients above this line were improved. Since the demand for psychotherapy far exceeds the supply of therapists available to do psychotherapy, this finding has considerable practical importance.

It is true that if on similar samples the *Es* scale were used for selection, some patients who would respond well to psychotherapy would not be chosen, but all those selected could accurately be predicted to show improvement.

Brief Report
Received July 26, 1955.

Reference

1. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

---

[1] An extended report of this study may be obtained without charge from Robert D. Wirt, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4672 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] Now at the University of Minnesota.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

The Generality of the Prediction of Self Reports

John J. Wittich[1]
DePauw University

The reverse-testing technique as proposed by Steinmetz (3, 4) is a procedure in which one person is asked to simulate the answers of another on a preference-type inventory. Although one use of the technique was to locate areas of misunderstanding or of differences in an individual marriage or parental counseling situation, some work has been reported in which groups utilized the cross administration principle (1, 2). Since little or no study had been devoted to the generality of the ability to predict or to the correlates of "understanding," a preliminary study was made (5).

Preliminary Study

With the limitations imposed by the procedure itself, the purpose of the previous investigation was: (*a*) to determine the degree of generality in the ability to predict the responses of others on a preference-type inventory, (*b*) to ascertain the relationship of the adjustment of one individual to his ability to predict the responses of others, and (*c*) to determine the relationship of the adjustment of a person to the success with which he is predicted.

In the study, each of 106 students chose two acquaintances to predict, item by item, on the Bell Adjustment Inventory. The predictions and the students' original responses were treated to yield adjustment scores for each predicting and predicted subject and two sets of prediction or "understanding" scores. The results of the intercorrelations of the distributions are tabulated in Table 1.

Pertinent partial correlations were also presented, and the following conclusions were drawn:

1. There is evidence that the ability to predict the responses of others may be regarded as a trait.
2. There is some evidence that the adjustment of the predictor is related to his ability to predict, but there is also evidence of an artifactual relationship.
3. There is some evidence that the adjustment of the person predicted is related to the ease with which he can be predicted, but there is also evidence of an artifactual relationship.

As a result of the preliminary study, the following suggestions for a future investigation were made: (*a*) the design of the experiment should not provide for the students to choose acquaintances to predict; (*b*) the design of the experiment should be such that the subjects will be required to predict several specific people with whom they have had time to become acquainted; and (*c*) the design should eliminate items on the instrument used which lack face validity. The present study meets these requirements.

The Problem

The primary objectives of the present experiment are (*a*) to test generality with a number of prediction, or understanding, tasks; (*b*) to re-examine with a new design the relation between adjustment and understanding; and (*c*) to investigate the relation between adjustment and the capacity to be understood.

Table 1

Intercorrelations of Five Variables
($N = 106$)

Variables	2	3	4	5
1. Total-Score-Student	.47	.37	.20	.08
2. Student-for-A		.48	.44	.16
3. Student-for-B			.04	.46
4. Total-Score-A				.25
5. Total-Score-B				

Note.—The standard error of an *r* of .00 is .09.

---

[1] This study is condensed from a section of a doctoral dissertation submitted to Stanford University in 1951. The author appreciates the assistance afforded by L. G. Humphreys, who was then on the Stanford faculty.

The Experiment

Subjects. The subjects were 42 enlisted men in the United States Air Force who had previously exhibited college-level ability on standardized batteries of tests. These subjects were not living together in barracks in the usual propinquity of military life, but they had been working together in closely-knit work units daily for at least four months. The range in age was from 19 to 33 years. These subjects were assigned to groups of six members each (within their own work units) and were required to predict and to be predicted by the other five group members.

Instrument. A revised and shortened Bell Adjustment Inventory was employed in this experiment since it was found that many items in the standard inventory, particularly those referring to health adjustment, were inappropriate for servicemen. The total number of items retained was 100.

Another important change entailed the elimination of the "?" as an alternative response. From the preliminary experiment it appeared that the effect of the "?" response for prediction may have been to increase the error involved.

It should be expected that the reliability of this revised instrument will be lower than that of the standard inventory since the revision has fewer items. The analysis of variance to be described later in this experiment will indicate that minimum reliability was obtained.

Directions. On the first day of the experiment, the airmen were given assurance that the self reports would not be used against them. After the self reports had been obtained from the subjects, answer sheets for the predicted responses were given them. Each predictor was instructed, in each case, to attempt to duplicate the answers of another member of his work group by responding to each item as he thought this subject would respond. Subsequently, each subject was required to attempt to duplicate the responses of every other subject in his group. Each airman, therefore, contributed to the experiment one self report and five prediction reports.

Scoring. Two types of scores are considered in the present experiment:

1. The adjustment scores were obtained by comparing the typical scoring key of the revised inventory with the observed responses of the airmen on their self reports.

2. The prediction scores were obtained by comparing, item by item, the predicted responses with the self report of the predicted subject.

Treatment of the scores. An analysis of the variance of the prediction scores was the test of generality, but the correlational method was applied to explore the relationships between adjustment, understanding, and the capacity to be understood. Three distributions were provided for statistical analysis:

1. The adjustment scores were obtained as described above.

2. The understanding distribution was obtained by averaging the prediction scores each airman obtained as he predicted five subjects.

3. The capacity to be understood was established by averaging the five prediction scores which were obtained for each predicted subject. This provided a distribution of mean scores.

Results and Discussion

Generality of the ability to predict. Table 2 contains the variance estimates which are considered appropriate for the purposes of the investigation.

It will be noted that the variance between groups has been computed. This was done so that the influence of the grouping may be eliminated from the F ratios under consideration. The table is broken down into the following sources of variance: (*a*) the variance between predictors, (*b*) the variance between predictees, and (*c*) a residual or interaction variance. The variance within the

Table 2

Variance Table for Scores Categorized in Terms of Predictors and Predictees

Sources	Sums of squares	*df*	Variance estimate
Between Groups	2,386	6	297.66
Between Predictors	4,683	35	133.80
Between Predictees	8,492	35	242.63
Residual (Interaction)	3,680	133	29.02
Within Predictors	12,352	168	73.52
Within Predictees	8,543	168	50.85
Total	19,421	209	

Predictor F ratio = 133.80/29.02 = 4.61; $p < .001$.
Predictee F ratio = 242.63/29.02 = 8.36; $p < .001$.

scores of predictors and the variance within the scores of predictees are also used.

In order to investigate the ability to predict, the variance between predictors was employed as the numerator and the residual or interaction variance was employed as the denominator. The resultant *F* ratio was significant beyond the .001 level.

It is possible to express the magnitude of the generality of the ability to predict in more than one frame of reference. Using the variance between predictors and the residual variance as the basis for computation, the average correlation between predictors was found to be .42. This coefficient, representing an intraclass correlation, may be interpreted as expressing the magnitude of the generality which might be expected if all the predictors predict subjects of equal predictability. In order to obtain generality of this magnitude, control of the predicted subjects would be necessary.

It is also legitimate in this situation to utilize the variance within predictors in combination with variance between predictors to compute an average coefficient of generality. When this is done the coefficient is equal to .14. These results are of the magnitude which would be expected if the experimental arrangements were such that the predictors were all predicting *different* subjects.

The level of significance of the *F* ratios indicates that the variance of the scores by predictors must be regarded as the result of nonchance differences between predictors in the ability to predict. There is generality in the ability to predict the responses of others, though the amount is of little practical significance unless there is control of predictees. The low average correlation of .42 between predictors under the more controlled condition indicates that any assumptions concerning the understanding ability of a subject on the basis of one or two predictions is subject to great error.

Generality of the capacity to be predicted. In order to investigate the capacity to be predicted, the variance between predicted individuals was used as the numerator and the residual or interaction term as the denominator. The resultant *F* ratio was significant beyond the .001 level. When the variance terms employed in the above *F* ratio were used to express the average correlation between predictees, the coefficient of generality was .59. Since this coefficient was derived from the variance between predicted individuals and the residual variance after the influence of group and predictor variance were removed, it is legitimate to interpret it as an expression of the expected results when all the predictees are predicted by predictors of the same ability to predict.

When the variance within predictees was employed with the variance between predictees to compute the average coefficient of generality, the resultant coefficient was .43.

These results, like those which concern the generality of the ability to predict, must be interpreted only in the light of the amount of restriction placed on the variance estimates. There can be little doubt that there is some evidence for nonchance differences between predictees in the capacity to be predicted, regardless of the variances used in the computation. The average coefficients of .59 and .43 indicate that the capacity to be predicted may be considered a trait.

Coefficients of correlation. Zero-order correlation coefficients with relation to the adjustment, understanding, and understood distributions were computed.

A zero-order correlation coefficient of .11 between the adjustment of the airman and his ability to understand others provides no evidence for a relation between adjustment and understanding. The correlation of .05 between understanding and being understood indicates that understanding, whatever its nature, is independent both of the predictor's adjustment and his capacity for being understood.

The correlation coefficient of .73 between the airmen's adjustment scores and the mean of the understanding scores which group members attained using them as subjects, merits extended consideration. It would appear from this coefficient that the better the adjustment of the predictee, the more easily he is understood. The viewpoint might be advanced that the predictors tend to predict adjusted responses. Then, if the person for whom they are attempting to duplicate responses is adjusted, the understood score will be high. If, however, the subject is maladjusted, the adjustment score will be low as will the level of understanding. However, an inspection of the distributions of prediction scores, scored for adjustment, found the means of these distributions to be about the same as the means of the actual adjustment scores of the subjects. The scatter diagram of the adjustment-understood relationship showed that, while the great majority of deviations from a roughly estimated best-fit line were instances where the predictee was misunderstood and well adjusted, there were no instances where the predictee was well understood and maladjusted. Support must be given, therefore, to a general thesis that the maladjusted *are* difficult to predict. The reason they are not understood is not because predictors refuse to predict maladjusted responses.

Another view of this finding may be found in the logical difference between predicting for an adjusted and a maladjusted individual. A predictor who believes he is predicting an adjusted subject marks the prediction report with the obvious adjusted answers. If he believes his subject is maladjusted, he has the task of determining which of the areas are maladjusted, and which items represent those areas. This is evidently a difficult task.

Conclusions

All conclusions are made in the frame of reference of the present study.

1. The ability to predict the responses of others may be regarded as a trait.
2. The capacity to be predicted by others may be regarded as a trait.
3. There is a positive relationship between the adjustment of a subject and the success with which others understand him.

Summary

The primary objectives of the experiment are (*a*) to determine the degree of generality in the ability to predict the responses of others on a preference-type inventory, (*b*) to ascertain the relationship of the adjustment of one individual to his ability to predict the responses of others, and (*c*) to determine the relationship of the adjustment of a person to the success with which he is predicted. A preliminary study had pointed out the necessity for an experimental design involving several prediction tasks for each subject.

A revised and shortened Bell Adjustment Inventory was employed as 42 subjects, arranged into six-man work groups, were required to predict and to be predicted by the other five group members.

Adjustment scores were obtained for each subject by utilizing the typical scoring key. Prediction scores were obtained by comparing, item by item, the predicted responses with the self report of the predicted subjects. These latter scores could be organized both in terms of the predicted scores which each subject obtained for others, and also by the predicted scores that were obtained by others for each subject.

An analysis of variance revealed that the capacity to be understood provides more substantial evidence for generality than the ability to understand, but that there is evidence for both. Intercorrelations between the adjustment of the subjects, their ability to understand others, and their capacity to be understood resulted in only one marked relationship, that between the airmen's adjustment scores and the mean of the understanding scores which group members attained using them as subjects. The thesis that the maladjusted are difficult to predict is supported.

It may be said in the frame of reference of the present study that (*a*) the ability to predict the responses of others may be regarded as a trait, (*b*) the capacity to be predicted by others may be regarded as a trait, and (*c*) there is a positive relationship between the adjustment of a subject and the success with which others understand him.

Received March 29, 1955.

References

1. Fiedler, F. A method for objective quantification of certain countertransference attitudes. *J. clin. Psychol.*, 1951, 7, 101–107.
2. Jessen, M. S. Parent-child cooperation in the counseling process. Unpublished doctor's dissertation, Stanford Univer., 1950.
3. Steinmetz, H. C. Directive psychotherapy vs. psychological understanding. *J. clin. Psychol.*, 1945, **1**, 331–335.
4. Steinmetz, H. C. Selecting personnel workers. *Educ. psychol. Measmt*, 1947, 7, 37–43.
5. Wittich, J. J. The generality of the prediction of self-reports. Unpublished doctor's dissertation, Stanford Univer., 1951.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Empathy, Similarity, and Self-Satisfaction[1]

Howard M. Halpern
Bronx VA Hospital

The recent review by Taft (7) of studies dealing with the ability to judge people reflects the current extensive interest and experimentation in this area. One problem that has attracted particular attention is the effect of similarity between predictor and predictee on the accuracy of predictions.

The issue is not a new one. In their experimental work on personality judgment, Wolf and Murray (8) noted that subjects were most accurate in predicting about people whose average ratings were most similar to their own and least accurate in making estimations about those whose average ratings were least similar. They thus raised a problem that is of utmost importance in the study of the empathic process. Do people empathize better with others who are relatively similar to themselves? If so, what is the role played by projection or attribution?[2] Wolf and Murray were inclined to believe that projection did not account for their result. They concluded, "The best explanation seems to be that man can only understand what he has already experienced. One might hazard the statement that without empathy a man cannot make an accurate diagnosis and he can best empathize with those whose responses resemble his own" (8, p. 363).

Bender and Hastorf (2) proposed an operational definition of projection as the relation between a subject's self-ratings and the ratings he attributes to others. From this they developed what they called a "refined empathy score" (3) by subtracting a subject's projection score from his total predictive accuracy. They felt that this refined score was a better measure of the subject's empathic ability. They noted also that while the raw empathy score correlated significantly with similarity, the refined score did not. Nevertheless, they found cause for dissatisfaction with their method of attempting to tease out projection (or attribution) from over-all predictive accuracy. They stated, "We suggest the need for a methodology for differentiating between projection and empathy in these prediction tasks. The next step would seem to be to obtain predictions from individuals for a number of their associates; these associates should differ in the amount of their similarity with the predictor. These data could be used to determine more closely the relationship between similarity, projection, and empathy" (6, p. 576).

One major aim of this study is the exploration of the relationship between similarity and empathy.

A second area of concern is the determination of the relationship between a person's self-satisfaction or self-dissatisfaction in a given personality area and his ability to predict accurately in that area.

Procedure

Subjects

The subjects were 38 female student nurses at the Adelphi College School of Nursing. For at least two years they had been in close association with the other members of their group in a variety of academic, social, and professional situations. The college faculty, employing sociometric methods, had divided the students into four "maximally harmonious" groups. Each group served a different hospital unit.

---

[1] This experiment was part of a doctoral research study at Teachers College, Columbia University (5). Grateful acknowledgments are extended to the author's dissertation committee, Professors Paul Eiserer, Edward J. Shoben, and Herbert Solomon. Dr. Gerald Bauman of the New York City Domestic Relations Court and Dr. Jacob Cohen, Chief, New York Training Unit, Veterans Administration, aided with the design and statistical treatment of the data. Dean Margaret T. Shay of the Adelphi College School of Nursing graciously made her students available as subjects.

[2] The term attribution rather than projection is preferred to avoid confusion with the psychoanalytic meaning of projection which refers only to the imputing of unconscious, unacceptable ideas to others. *Attribution* does not differentiate between consciously or unconsciously imputed characteristics.

Instrument

With the aid of judges and a priori criteria, an instrument was devised consisting of 80 items from the Guilford-Martin Inventory of Factors GAMIN (4)[3] that were considered most suitable for use in a predictive empathy test. An equal number of items was selected from each of the five factors.

Two further changes of the GAMIN inventory were made, one by omission and one by addition. The opportunity to respond to an item with a question mark, if a "Yes" or "No" answer could not easily be given, was eliminated, thus forcing the subjects to respond dichotomously. The second change involved asking the subjects to indicate whether they were "Pleased" or "Dissatisfied" with themselves in the personality area tapped by each item by encircling a "P" or a "D" after answering the item.

Administration

Each group was seen twice. At the first session the subjects were asked to rate themselves on the revised GAMIN inventory. Before the groups were seen again, each subject's similarity to each of the other members of her group was determined by tabulating the number of "Yes" or "No" items that they answered in the same way. Then, the two group members most Similar and most Dissimilar to each subject, plus a member between these extremes, were selected to function as predictees or referents for that subject.

When the subjects were seen two weeks later they were instructed to predict how each of their five referents responded to the inventory during the first session.

Hypotheses

1. Subjects predict with greater accuracy about predictees who are similar to them than about those who are dissimilar to them.

2. Subjects predict with greater accuracy on concordant items (items which they have answered in the same way as their predictees) than on nonconcordant items (items which they have answered differently from their predictees).

3. On nonconcordant items there is greater accuracy in the predictions made about similar referents than dissimilar referents.

4. Ability to predict accurately on nonconcordant items correlates positively with over-all predictive accuracy.

5. On items where subjects have indicated that they are dissatisfied with their self-ratings they do not predict as accurately as on items where they have indicated that they are pleased.

Results

1. A *t* ratio indicated that predictions for Similar referents were more accurate than for Dissimilar referents at a level of significance beyond .001 (*t* was 4.25 for 38 subjects). Furthermore, a highly significant 30 of the 38 subjects were better predictors when putting themselves in the place of those they resembled. When *mean similarity scores* (the average number of items each subject answered in concordance with each of her five referents on the self-ratings) were correlated with *mean predictive accuracy* (the average number of correct predictions each subject made about her five referents) the resulting *r* was a significant .84.

2. The second hypothesis stated that the subjects predict with greater accuracy on concordant items than on nonconcordant items. A *t* ratio of 12.81 supported this hypothesis.

3. On nonconcordant items, subjects did not predict significantly more accurately for Similar than for Dissimilar predictees (*t* was 1.19).

4. A nonsignificant correlation of .05 was found between a subject's ability to predict accurately on nonconcordant items and her over-all predictive ability.

5. The results of a *t* test indicated that subjects predicted with greater accuracy on

---

[3] (G) general pressure for overt activity, (A) ascendency as opposed to submission, (M) masculinity as opposed to femininity in attitudes and interests, (I) inferiority feelings, and (N) neurotic tendencies.

"Pleased" items than "Dissatisfied" items at beyond the .001 level of significance.[4]

Discussion

The data reflect a clear positive relationship between an individual's similarity to an acquaintance and his ability to make accurate predictions about him. Furthermore, predictive accuracy is greater when the individuals involved resemble each other in the specific areas of prediction than when they differ.

Empathy or Attribution?

What accounts for this ability of individuals to prognosticate better about Similar as opposed to Dissimilar people and within shared rather than unshared areas of personality? Empathy has been operationally defined in terms of predictive accuracy, but it is apparent that the above result could be a function of the attribution of a subject's own traits to others rather than genuine sensitivity.

If, in nonconcordant areas, the subjects had predicted more efficiently about Similar rather than Dissimilar acquaintances this greater accuracy could only be a function of true sensitivity to similar people. In unshared areas, attribution can lead only to error and therefore could have been ruled out as a facilitator of more accurate prediction.

Such was not the case. Subjects did not predict significantly better for Similar as opposed to Dissimilar referents in these unshared areas. The increased over-all predictive ability found to exist when the predictor and predictee were similar was confined to the areas of their similarity. Once outside these concordant areas the predictors were just as much at sea when their total resemblance to their referent was relatively great as when it was minimal.

Because a person's ability to empathize with similar people lies strictly within the boundaries of their similarity and disappears outside of these boundaries it might seem, at first glance, that an attributive mechanism must be at work. This is not necessarily so. What may instead be indicated is that people cannot effectively predict about what they have not phenomenologically experienced. The close relationship between similarity and predictive skill might simply mean that there is a greater likelihood that a person would *recognize* feelings and patterns of behavior in others if he has known similar feelings and patterns of behavior in himself. When an individual is confronted with the emotions and actions of another person that are alien to his own experience, accurate recognition, and hence, accurate prediction, is little better than a chance matter.

The degree of similarity between two people may therefore reflect the overlapping of their capacities to empathize with each other. The rationale for the finding that even two grossly similar people cannot predict at better than chance accuracy in the few areas of dissimilarity would then be that they were attempting to empathize where little capacity for empathy existed.

Although the equation of similarity with capacity for empathy has rarely been so pointedly suggested by experimentation, it is not a new idea. In 1928 Adams wrote:

> . . . any experience or mental process in another organism can be inferred from structure, situation, history and behavior only when a similar experience or mental process is or has been invariably associated with similar structure, situation, history and behavior in oneself; and the probability of the inference will be proportional to the degree of similarity (1, p. 252).

Two alternative concepts, therefore, still remain as possibilities in explaining the finding that the greater predictive ability for similar people is restricted to the areas of their similarity. Either the concordant areas represent the presence of a capacity for mutual empathy, or are simply areas where attribution will inevitably get good results. The methodology of this study has failed to differentiate the relative weight of these two factors.

When the subjects' ability to predict accurately on nonconcordant items was compared with their over-all predictive accuracy, no relationship was found. The question arises, which is the truer measure of empathy? Is it the ability to feel most extensively into the largest number of people, or the ability to feel into those characteristics of others that differ from one's own?

---

[4] The complete data of this study may be obtained from University Microfilms, 313 No. First Street, Ann Arbor, Michigan. Refer to H. M. Halpern, Some factors involved in empathy, Publication No. 8677.

Both arguments can be persuasive. It must certainly be granted that a measure which clearly eliminates attribution seems closer to the spirit of the term empathy. Yet, as we have seen, because a person accurately enters into the frame of reference of a similar person we cannot assume attribution nor rule out empathic sensitivity. Further, it would seem to be doing just as great an injustice to the meaning of the term empathy to designate as good empathizers people who cannot feel extensively into the phenomenological field of others but who can somehow predict about them relatively better in areas where they differ.

Empathy and Self-Satisfaction

The subjects predicted with significantly greater accuracy on items where they had indicated self-satisfaction rather than dissatisfaction. One possible rationale for this finding is that in areas where a person is discontent about his own behavior, disorganizing anxiety may be aroused and distortive defenses may be mobilized. Both of these factors may function to cause aberrations in accurate interpersonal perceptions.

Who Is the Good Empathizer?

The close relationship that was found between similarity and empathy, particularly if similarity is considered a measure of the empathic potential existent in the situation, leads to inferences about the personality of the good empathizer. First of all, the phenomenological experiences of the good empathizer could not be drastically deviant from those of his reference group. He has, all things being equal, capacity for empathy with more people if he is near the center of the range on a given characteristic. Secondly, the wider his phenomenological experience, in terms of its breadth, fullness and richness, the more people will he be able to encompass, through similarity, in his empathic scope.

The relationship between self-satisfaction and empathy suggests that a person who is at home with most of his own behavior is likely to be a better empathizer than those who are largely dissatisfied with themselves.

Summary

The ability of 38 student nurses, in four groups, to predict each other's responses to a personality inventory was determined. This ability, which served as an operational definition of empathy, was found to be positively correlated with (*a*) the similarity of predictor and predictee, and (*b*) the predictor's satisfaction with her own behavior in the area of prediction.

The methodology of this study failed to resolve whether the ability to predict more accurately for similar predictees was a product of attribution or true sensitivity. However, it is proposed that the factor of similarity may not be an artifact of the predictive method, as other experimenters have suggested, but that similarity may be a vital part of the empathic process. It may be that people can most readily recognize in others what they have experienced, on some level, in themselves.

Received May 2, 1955.

References

1. Adams, D. K. The inference of mind. *Psychol. Rev.*, 1928, **35**, 235–252.
2. Bender, I. E., & Hastorf, A. N. The perception of persons: forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 556–561.
3. Bender, I. E., & Hastorf, A. N. On measuring generalized empathic ability (social sensitivity). *J. abnorm. soc. Psychol.*, 1953, **48**, 503–506.
4. Guilford, J. P., & Martin, H. G. *The Guilford-Martin inventory of factors GAMIN. Manual of directions and norms.* Beverly Hills, Calif.: Sheridan Supply Co., 1943.
5. Halpern, H. M. Some factors involved in empathy. Unpublished doctor's dissertation, Teachers College, Columbia Univer., 1954.
6. Hastorf, A. H., & Bender, I. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 574–576.
7. Taft, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.
8. Wolf, R., & Murray, H. A., An experiment in judging personalities. *J. Psychol.*, 1937, **3**, 345–365.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Cultural Symbolism: A Validity Study

Emanuel Starer
Veterans Administration Hospital, Coatesville, Pa.

Levy (1) reported that the results of his experiment with 62 normal children in the fifth grade of a public school did not support the hypothesis that ". . . certain environmental objects by virtue of their structural characteristics may be taken as being symbolic of the male and female genitalia" (1, p. 45).

The purpose of the present study is to investigate this same hypothesis that "where abstract objects are categorized by a given means as either male or female, those objects which may be described as elongated, pointed, or angular will be categorized as male; those objects which may be described as round and containing in nature will be categorized as female" (1, p. 43).

Method

The *S*s in this experiment consisted of 64 male psychotic patients and 48 female psychotic patients in varying degrees of remission. The age range of the *S*s in the male group and female group was 20–48 and 24–50, respectively. The primary diagnosis in 98 per cent of the psychotic population was schizophrenic reaction. Another population consisting of 30 student nurses with a mean age of 19 and considered to be clinically normal was also tested.

The *S*s were presented with ten cards containing figures which were presented in random order. Five of the figures were elongated or pointed and were assumed to be symbolic of male characteristics. Five figures were rounded and containing and symbolic of female characteristics.[1] Ten additional cards containing the names of five males (Jack, Thomas, Sam, James, and Harry) and five females (Mary, Betty, Irene, Sally, Barbara) were given to each subject. These were also arranged in a randomized manner. Each subject was seen individually for approximately ten minutes and was told to "place the names of these five men and women on these cards (containing the 5 male and 5 female figures) wherever you feel they belong." After this was done and the results recorded, the *S*s were asked to tell the examiner what each figure looked like to them.

Results

As in Levy's experiment the presented task requests a simple sorting problem. By calculating the expected probability of getting any given number of "correct" sorts and comparing it with the obtained frequencies by means

Table 1

Predicted and Observed Frequency Distributions for Any Given Number of Correct Matches in the Male Psychotic Population ($N = 64$)

Number of correct matches	p	Predicted frequency	Observed frequency
0	.001	.06	0
2	.10	6.40	1
4	.39	24.96	13
6	.39	24.96	18
8	.10	6.40	23
10	.001	.06	9

Table 2

Predicted and Observed Frequency Distributions for Any Given Number of Correct Matches in the Female Psychotic Population ($N = 48$)

Number of correct matches	p	Predicted frequency	Observed frequency
0	.001	.05	0
2	.10	4.80	0
4	.39	18.72	5
6	.39	18.72	21
8	.10	4.80	17
10	.001	.05	5

---

[1] Copies of the figures are available on request from the author.

Table 3

Predicted and Observed Frequency Distributions for Any Given Number of Correct Matches in the Student Nurse Population ($N = 30$)

Number of correct matches	p	Predicted frequency	Observed frequency
0	.001	.03	0
2	.10	3.00	1
4	.39	11.70	2
6	.39	11.70	8
8	.10	3.00	9
10	.001	.03	10

of chi square, it is possible to test the null hypothesis.[2] The results are presented in Tables 1, 2, and 3.

The chi square for Table 1 is 113.17 with three degrees of freedom. The result is significant beyond the .001 level of significance. Chi square for Table 2 with three degrees of freedom is 75.78 and is significant beyond the .001 level of significance. Chi square for Table 3 with three degrees of freedom is 94.73. This result is significant beyond the .001 level of significance.

Table 4 contains the most frequent associations to the presented figures given by the three groups of subjects.

Discussion

The results of this experiment are contrary to that obtained by Levy. It is possible that some of the discrepancy may be explained on the basis that distinct differences exist in the type of subject population. Again, in the present experiment, a different set of stimuli was employed. However, inspection of the two sets of stimuli suggests that they are substantially similar. Another factor is the possibility that Levy used group testing, although this is not absolutely clear from his report, and only individual testing was used in the present experiment.

Table 4

Most Frequent Associations to the Presented Figures in the Male and Female Psychotic Population and Student Nurse Population

Figure	Association
1	Pencil, bullet
2	Lips, mouth
3	Gun stock, or rifle
4	Side view of a canoe, stick, branch, belt
5	Pen, pencil
6	Hat, bowl
7	Hills, mountains, letter "m"
8	Bowl, cup
9	Cigarette
10	Circular form or frame, letter "o," ring

Before we can dismiss the existence of a cultural sexual symbolism, more extensive investigation is needed. Certainly, the results of the present experiment point to the strong possibility that there is such a phenomenon as a symbolism which is generally accepted in any particular culture.

An interesting observation made by the examiner, although statistical proof cannot as yet be presented, was that the patients who were relatively more confused and disorganized made many more errors in matching than the patients who seemed in fairly good states of remission. The observation raises the possibility that one source of illness is the fact that there never has been a complete acceptance on the part of the very disturbed patient of the generally prevailing symbolism in a given culture. Another possibility is that there has been a regression as a result of emotional illness to a level where cultural symbolism does not yet play an effective role for the personality. This appears to warrant further investigation.

Summary

The results of the present investigation appear to support the hypothesis that there may be a symbolism in any particular culture which is generally accepted by those who are functioning relatively effectively. The possibility is raised that the inability of certain individuals to adapt to and accept cultural symbolism results in emotional disturbance.

Received May 6, 1955.

Reference

1. Levy, L. H. Sexual symbolism: a validity study. *J. consult. Psychol.*, 1954, **18**, 43–46.

---

[2] In calculating χ^2, the two extreme categories at either end were combined because of the low predicted frequencies.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Temporal Acuity of Vision, Audition, and Touch in Psychogenic and Neurogenic Pathology[1]

Albert F. Ax
University of Washington

and William H. Colley
VA Mental Hygiene Clinic, Huntington, West Virginia

Critical fusion frequency, as Geldard (6, p. 89) suggests, may be considered as a measure of visual "temporal acuity." Temporal acuity is the ability to resolve two or more discrete stimuli appearing close together in time. As the time is decreased between recurring stimuli, the critical frequency is finally reached at which the separate stimuli lose their discreteness with a noticeable and reportable change in the quality of the sensation. For visual stimuli at frequencies above the critical flicker fusion point, the sensation is a steady light. Below the CFF frequency there is a sensation of flicker. The many studies of visual temporal acuity (CFF) have recently been reviewed by Hecht and Verrijp (8) and by Simonson and Brozek (14).

Temporal acuity for other sense modalities has not been so extensively studied. In discussing "auditory flutter," Geldard reports the phenomenon not to be comparable to visual critical flicker fusion, since "if the tonal cutoff is really effective the 'critical fusion point' is never reached . . . [it] is in no sense the 'fused' continuity of the kind found in vision above the critical frequency of flicker" (6, p. 137). It is clear, from the context, Geldard is oriented in this discussion to the nature of the sensation above the critical point as related to problems of neural persistence. If, on the other hand, the aspect of "temporal acuity" is considered, the investigation can be focused on whether subjects can consistently report the point at which the sensation of separate stimuli is lost (with rising frequency), or when the sensation of discreteness reappears with decreasing frequency.

Temporal acuity of the skin to tactual and electric stimuli has been largely neglected in the long discussions of the "vibratory sense" which Geldard (5) has thoroughly reviewed. He does, however, report a few findings on the lower frequency limit for the "vibratory sense," that is, the lowest frequency at which separate impacts are felt as a "vibration" or separate shocks as a "tingle." Knudsen (10) found the lower limit to be from 12.5 to 18.00 per second for shocks with an average of 15.3 for two observers. Kampik (9) and Brecher (4) believed it to be 16 cycles per second for tactual impacts.

These studies indicate that temporal acuity for the tactual sense can be identified. It is clearly recognized, however, that as with the auditory temporal acuity, when the stimulus frequency is above the critical value, the experience is quite different from the smooth, continuous sensation of the steady-appearing light when above the CFF frequency.

Since the various sense modalities produce a variety of experiences when stimulated at frequencies above the temporal acuity threshold, it is essential to define the temporal acuity threshold by the uniform referent of "loss of discreteness." Seeing, hearing, or feeling separate, discrete stimuli is a general experience, not dependent on the particular sense modality. Accordingly, in this experiment the temporal acuity threshold is defined as the mean frequency at which the experience of discreteness is lost with increasing frequencies and reappears with decreasing frequencies.

---

[1] From the Department of Psychiatry, University of Washington School of Medicine and the Veterans Administration Hospital, Seattle, Washington.

Critical fusion frequency of vision (CFF) has repeatedly been found to be lower in various pathological conditions of the central nervous system produced by toxins, anoxia, and physical damage to neural tissue. Battersby, Bender, and Teuber (2) recently reported on a study of CFF in patients with long-standing lesions of either the frontal or occipital lobes. On the basis of their findings, they concluded that transient depressions in visual FF may occur after frontal lobe lesions, but that injuries to higher *visual* pathways are necessary to produce a lasting reduction in CFF. They further hypothesized that transient reductions of CFF might appear after cortical ablation of other nonvisual areas involving, for example, the temporal or parietal lobes.

If this view is correct, the usefulness of visual CFF alone, as a device for localizing brain pathology, would be limited to statements of the absence or presence of an acute or new pathology of unknown position in the cortex, or the presence of an old lesion in higher visual pathways. In any given patient, which of these is the case could hardly be specified with much confidence on the basis of visual CFF data alone. Were it possible to demonstrate that lasting depressions of temporal acuity in other sense modalities follow involvement of their higher pathways and projection areas, localization of old lesions could be more precise. As in vision, a new or acute pathological condition of any lobe might still result in depressed temporal acuity in other modalities, but the data obtained from their simultaneous study might reveal different degrees of depression and make possible localization of pathology while it is still new or in an acute phase.

Halstead (7) has suggested a central mechanism of the cortex which limits the resolving power of the primate visual system. On the other hand, there is the possibility that each of the senses has its own integrating mechanism or center governing its temporal resolving power which would permit the temporal acuity of the various senses to be independent of each other, so long as lesions are not disruptive of the general physiological economy of the total cortex. Where a new or acute disturbance affects the total cortex, temporal acuity of all modalities may be depressed, giving the appearance of a central fusion mechanism.

The present investigation seeks to determine, first, whether temporal acuity can be reliably measured for auditory and tactual sense modalities and secondly, whether this extension to other sense modalities will provide an increase in power for the diagnosis and localization of brain pathology. A third question as to whether the fusion mechanism is a general function of the cortex as a whole or whether it is specific to the sense modality will be explored, but it is unlikely to be clearly answered by this study.

Subjects

All subjects were male patients at the Veterans Administration Hospital, Seattle, Washington. Acceptability was contingent upon being ambulatory, capable of understanding and following instructions, and remaininig hospitalized for a period permitting careful medical evaluation.

The neuropathology group consisted of 22 patients. The criteria for neuropathology were positive neurological signs, abnormal electroencephalograms and, where available, positive X ray and neurosurgeons' reports. Except as implied above, the subjects were unselected as to extent or location of pathology. Final diagnoses revealed three patients with frontal lobe lesions, two with parietal lobe lesions, two with temporal lobe lesions, one with a cerebellar lesion, three with subcortical lesions, and eight with generalized cortical pathology. Of the remaining three, lesions were found in the occipital, temporal, and parietal lobes of one, in the temporal and frontal lobes of the second, and in the temporal and occipital lobes of the third. Anticonvulsant and/or sedative drugs were being administered to 10 of the group on the date of testing. Ages ranged from 22 to 62 years, with a mean age of 40.4 years. Cerebrovascular accident patients, by virtue of their generally higher age and ease of diagnosis, were eliminated from consideration.

The psychopathology group of 21 consisted of five patients initially suspected, but not subsequently diagnosed as having brain pathology, and 16 patients not suspected or found to have neuropathology during their medical evaluation. Selected at random from the psychiatric wards, 11 were classified as psychoneurotics (three conversion reactions and eight other or mixed subtypes), seven as psychotic (two psychotic depressions and five schizophrenic reactions), and three as character disorders. Anticonvulsant and/or sedative drugs were being administered to four of the group when tested. Ages ranged from 25 to 62 years, with a mean age of 37.7 years.

Apparatus

The intermittent stimuli were produced by a General Radio Strobotac (type 631-B) equipped with amplifier, speaker, tactual stimulator, and neon tube. This oscillator had two pulse frequency ranges of 10 to 60 cycles per second for the low range, and 40 to 420 cycles per second for the high range.

For the visual stimuli, a type NE 40 neon tube was plugged directly into the "strobolux" output jack, which provided a brief, high voltage pulse of less than 30 microseconds' duration. The NE 40 tube was painted black except for a ½-inch circular window directly over the glow plate. The tube was viewed through a light-tight black box, the light source being 40 inches from the subject's eyes.

Tactual stimuli were produced by a bamboo stylus projecting vertically through an adjustable washer mounted on a horizontal surface of convenient height for the subject to rest his hand. The stylus was operated by a relay modified to have approximately one millimeter stroke. Since the touch vibrator could not be made completely noiseless, a masking sound was produced (during touch only) by a 60-cycle hum through earphones from the 6.3-a.c. heater supply of the Strobotac. The hole in the washer on which the subject rested his finger was very small (about ⅛ inch in diameter) which prevented even large variations in finger pressure to dampen or impede the stroke of the stylus. In all cases, the stylus was felt to continue to vibrate at all frequencies used.

Auditory and electric stimuli were provided by a condensor-coupled cathode follower amplifier. The rectified output of this amplifier was switched either to a 3-inch speaker via a 25,000-ohm output transformer, or to two ¼-inch silver electrodes attached to the volar surfaces of two fingers on one hand with electrode jelly and tape. The sound produced was a clean-cut "tick." The shock was also a clean-cut, discrete "jolt." A volume control provided an intensity range from below threshold to an intensity well above threshold for all subjects for both auditory and shock stimulation.

Procedure

The temporal acuity thresholds for the four types of stimulation were determined in the following order: (*a*) vision, (*b*) audition, (*c*) touch, and (*d*) electric. Twenty alternate up and down trials were given for each modality.

The first starting frequency was 10 cycles per second and thereafter randomly varied from 10 to 20 cycles per second above and below the previous threshold. Rate of change was one cycle per second per second.

Instructions were similar for all tests. For example, the auditory temporal acuity threshold (ATA) instructions were: "Can you hear this ticking sound? You can hear each tick separately, but as I speed it up they will reach a rate so fast you cannot hear the separate ticks, but rather, the sound will become a hum or buzz. Say 'now' when you lose track of the separate ticks." (The *E* demonstrated.) "Now as I slow it down, the separate ticks will come back. Tell me when you hear the separate ticks again."

The instructions for visual (VTA), touch (TTA), and electric (ETA) temporal acuity frequency thresholds were similar. In all cases the disappearance and reappearance of the intermittent aspect of the phenomenon were indicated as the focus of attention, although the natures of the sensations produced by frequencies above the thresholds were named, such as hum or buzz for audition, vibration for touch, and "tingle" for electric. The shock current was set at 1.5 times the mean of three

Table 1

Temporal Acuity Values for Four Stimulus Modalities
Average of Up Plus Down Trials

Group	Vision $\bar{X}$	Vision S_x^2	Audition $\bar{X}$	Audition S_x^2	Touch $\bar{X}$	Touch S_x^2	Electric $\bar{X}$	Electric S_x^2
PP	31.30	21.51	68.63	1,041.33	34.62	31.29	33.73	29.56
NP	28.59	20.88	53.56	853.19	32.20	78.14	30.94	48.71
Diff.	2.71		15.07		2.42		2.79	
F						2.50		1.65
t	2.73		2.27		1.08		2.07	
Prob.	<.01		<.05		>.10	.02	<.05	.10

rising threshold determinations. Because of sensory adaption or rising resistance, it was found necessary to redetermine the threshold every fifth trial. Most subjects reported this shock intensity of 150 per cent of threshold to be clearly perceptible, but not uncomfortable.

Except for the shocks, the stimuli were set at a constant intensity, as is customary in CFF studies. It seems likely that greater reliability of temporal acuity measurement could be made if the stimulus intensity were adjusted to some standard decibel level above the subject's threshold.

Results

The customary score for flicker fusion is the mean frequency for the up and down trials averaged together. In Table 1 these scores are tabulated for each of the four stimulus modalities. On the first line are the means and variances (computed with $N-1$) for the psychopathology (PP) group of 21, and just below, for the neuropathology (NP) group of 22. The mean on visual temporal acuity (VTA) for the psychopathology group is 31.30, and for the neuropathology group, 28.59 cycles per second. The average difference of 2.71 cycles per second has a t value of 2.73 and a null probability of less than 1 per cent. In the next column for auditory temporal acuity (ATA), the difference is 15.07 cycles per second, but because of the extremely large variances, the null significance level falls between 1 per cent and 5 per cent. The touch temporal acuity (TTA) does not differ significantly for the two groups, although the mean is lower for the neuropathology group. The variance is 2.5 times as large for this group, with an F ratio significant at the 2 per cent level. The mean electrical temporal acuity (ETA) for the neuropathology group is 2.79 cycles per second lower, a difference significant at the .05 level. Here again, the variance is considerably larger for the pathology group although the F ratio is not significant.

In order to check whether the up and down

Table 2

Errors of Habituation for Four Stimulus Modalities
Average of Up Minus Down Trials

Group	Vision $\bar{X}$	Vision S_x^2	Audition $\bar{X}$	Audition S_x^2	Touch $\bar{X}$	Touch S_x^2	Electric $\bar{X}$	Electric S_x^2
PP	−1.55	3.64	13.47	397.94	1.11	57.41	.21	27.42
NP	.13	4.79	17.53	599.01	6.89	60.44	4.12	49.55
Diff.	−1.68		−4.06		−5.78		−3.91	
F								2.03
t	3.80		.26		3.59		2.81	
Prob.	<.01		>.10		<.01		<.01	.06

Table 3

Average of Up Trials

	Vision			Audition			Touch			Electric		
Group	$\bar{X}$	S_x^2	S_i^2	$\bar{X}$	S_x^2	S_i^2	$\bar{X}$	S_x^2	S_i^2	$\bar{X}$	S_x^2	S_i^2
PP	30.52	19.72	1.20	75.36	1,644.82	110.95	35.17	34.71	13.30	34.84	27.21	15.89
NP	28.66	16.40	1.57	62.32	1,443.35	383.45	35.65	95.45	20.75	33.00	63.20	30.70
Diff.	1.86			13.04			−.52			.84		
F								2.76			2.33	
t	2.03			1.54			.27			.61		
Prob.	<.05			<.10			>.10	.05		>.10	.05	

trials might be contributing differentially to these group differences, the up and down trials are averaged separately and posted in Tables 3 and 4, respectively. It is immediately apparent that the down trials are contributing most of the difference, since the down trial means are significantly lower for the neuropathology group in all stimulus modalities than are the up trials. Only the VTA has a significant difference on the up trials at the .05 level.

In order to isolate this factor which is contributing to the down trials over and above the temporal acuity level itself, as measured by the mean of the ups plus downs, the differences were taken between the ups and downs which are listed in Table 2. All but vision are heavily loaded with what may be described as an "overshoot" factor. This "overshoot" tendency has been described by Underwood (15) as "error of habituation," and when negative, as the "error of expectation." The neuropathology group tends to "overshoot" their temporal acuity level; that is, they go higher on the up trials and lower on the down trials. In the visual modality the psychogenic group has a mean "error of expectation" of −1.55 while the neurogenic group has an almost zero mean error of .13.

Each subject's individual variances from trial to trial are tabulated in the columns of Tables 3 and 4 headed as "S_i^2 with the up and down trials separate. With one exception, that is, the auditory down trials, these individual variability scores average higher for the neuropathology group. However, the variability scores themselves have such large variances in both groups as to make the t tests of the differences between groups not significant.

The correlations in Table 5 show an interesting trend. On the measures of temporal acuity, the neuropathology group have very small correlations between the four modalities, all being insignificant except that between touch and electric. This high correlation between the touch and electric stimulus modalities suggests these two modalities apparently involve the same sensory projection areas. In contrast to the low average inter-

Table 4

Average of Down Trials

	Vision			Audition			Touch			Electric		
Group	$\bar{X}$	S_x^2	S_i^2	$\bar{X}$	S_x^2	S_i^2	$\bar{X}$	S_x^2	S_i^2	$\bar{X}$	S_x^2	S_i
PP	32.07	27.20	1.41	61.89	1,130.32	106.64	34.06	55.52	16.19	33.63	41.64	19.50
NP	28.53	27.26	1.75	44.79	563.66	79.13	28.76	90.60	22.53	28.88	59.39	29.46
Diff.	3.54			17.10			5.30			4.75		
F					2.00							
t	3.15			2.72			2.88			3.10		
Prob.	<.01			<.01	.06		<.01			<.01		

Table 5

Intercorrelations* Between Temporal Acuity Measures of Four Stimulus Modalities

Variables	Psychopathology $N = 21$				Neuropathology $N = 22$				Combined groups $N = 43$
	U+D	U−D	Up	Down	U+D	U−D	Up	Down	Down
V A	56	28	56	53	00	31	01	09	40
V T	57	32	57	53	09	56	09	15	38
V E	55	15	75	46	02	49	00	13	35
A T	57	54	54	60	26	76	42	24	46
A E	57	47	55	56	41	44	51	37	50
T E	75	85	75	78	55	85	56	67	74
Mean *r* (via *z*′)	60	48	63	59	24	60	28	29	48

* Decimal points omitted.

correlation of .24 for the neuropathology group, the TA modalities are well correlated for the psychopathology group, averaging .60. This finding of contrasting intercorrelations for the two groups suggests the possibility that specificity of the brain lesion is revealed by specificity of temporal acuity in that sensory modality mediated by the particular brain area damaged. The intercorrelations for the up minus down trials appear to show an opposite trend. For these "habituation" scores, the neuropathology group have an average correlation of .60 contrasted to .48 for the psychopathology group. "Habituation" as an attitude or personality characteristic would be expected to be a generalized effect and not specific to the modality. The various modalities may vary, of course, in their effectiveness as a reliable measure of the "habituation" characteristic.

An attempt was made to examine more directly the relationship between the locus of lesion and the modality most affected by grouping the subjects into frontal, parietal, temporal, and occipital lobe lesions. No relationship, however, was apparent, which may be a function of the small size of the subgroups and the lack of certainty of the precise localization for many.

There were five cases with known unilateral lesions. These were tested on both hands for touch and electricity. There was no trend for the temporal acuity threshold to be related to laterality of lesion.

Finally, the effect of sedatives or anticonvulsants, which some of the patients were taking, was related to temporal acuity. Of the neuropathology group, 10 were receiving drugs and 12 were not. The medicated group had *higher* temporal acuity thresholds on all sense modalities. The "error of habituation" was *less* for the medicated group. Since only four (19%) of the psychopathology group were receiving drugs, in contrast to 10 (45%) of the neuropathology group, these findings indicate that not only could the drugs *not* account for our general findings of lower temporal acuity thresholds and higher habituation errors for the neuropathology group, but, rather, the drug effect may have attenuated the differences found.

Conclusions

1. These results indicate that temporal acuity thresholds for audition and touch can be measured with sufficient reliability to be useful for the diagnosis of cerebral pathology. The best single indicator is the vision "error of habituation" which correctly diagnosed 33 cases (76%), using the median as the cut-off point. Vision down trials were almost as good. When the touch variability scores were combined with these two, all three being equally weighted, 35 cases (81%) were correctly classified with three psychopathology and five neuropathology cases being misclassified. Hence, for these samples there is gain in

diagnostic power by combining several sense modality acuity thresholds and "errors of habituation" into a battery.

2. The question as to whether fusion frequency is a generalized function of the brain as a whole, or specific to each sense modality, remains obscure. While the lower intervariable correlations for the pathology group argue for specificity, our available samples with specific localized lesions are too small to draw any satisfactory conclusion.

3. The finding that the "error of habituation" score is reflecting a characteristic highly diagnostic of cerebral pathology suggests that other psychophysical measures, not involving temporal acuity, should be investigated as measures of this perseverative tendency often seen in patients with cerebral pathology.

Received May 2, 1955.

References

1. Battersby, W. S. The regional gradient of critical flicker frequency after frontal or occipital lobe injury. *J. exp. Psychol.*, 1951, **42**, 59–68.
2. Battersby, W. S., Bender, M. B., & Teuber, H. L. Effects of total light flux on critical flicker frequency after frontal lobe lesion. *J. exp. Psychol.*, 1951, **42**, 135–142.
3. Bishop, G. H. Responses to electrical stimulation of single sensory units of the skin. *J. Neurophysiol.*, 1943, **6**, 361–382.
4. Brecher, G. A. Die Entstehung und biologische Bedeutung der subjectiven Zeitenheit,—des Momentes. *Z. vergl. Physiol.*, 1933, **18**, 204–243.
5. Geldard, F. A. The perception of mechanical vibration: III. The frequency function. *J. gen. Psychol.*, 1940, **22**, 281–289.
6. Geldard, F. A. *The human senses.* New York: John Wiley & Sons, 1953.
7. Halstead, W. C. *Brain and intelligence.* Chicago: Univer. of Chicago Press, 1947.
8. Hecht, S., & Verrijp, C. D. Intermittent stimulation by light. IV. A theoretical interpretation of the quantitative data of flicker. *J. gen. Physiol.*, 1933, **17**, 269–282.
9. Kampik, A. Experimentelle Untersuchungen über die praktische Leistungsfähigkeit der Vibrationsempfindungen. *Arch. f. d. ges. Psychol.*, 1930, **76**, 3–70.
10. Knudsen, V. O. Hearing with the sense of touch. *J. gen. Psychol.*, 1928, **1**, 320–352.
11. Miller, G. A., & Taylor, W. G. The perception of repeated bursts of noise. *J. acoust. Soc. Amer.*, 1948, **20**, 171–182.
12. Phillips, G. Perception of flicker in lesions of the visual pathways. *Brain,* 1933, **56**, 464–478.
13. Ricciuti, H. N. A comparison of critical flicker frequency in psychotics, psychoneurotics, and normals. *Amer. Psychologist,* 1948, **3**, 276–277. (Abstract)
14. Simonson, E., & Brozek, J. Flicker fusion frequency: background and applications. *Physiol. Rev.*, 1952, **32**, 349–378.
15. Underwood, B. J. *Experimental psychology.* New York: Appleton, 1949.
16. Werner, H., & Thuma, B. D. Critical flicker frequency in children with brain injury. *Amer. J. Psychol.*, 1942, **55**, 394–399.
17. Young, K. M. Critical flicker frequency. In F. A. Mettler (Ed.), *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Social Desirability and Q Sorts

Allen L. Edwards
The University of Washington

In Q technique, as used in clinical and personality research, subjects are often asked to describe themselves in terms of a set of personality items. If these items differ with respect to their social desirability scale values, then social desirability may function as an uncontrolled variable in the Q sorts of the subjects. The potential influence of social desirability on the statistical analysis of Q sorts has been described in some detail by Edwards and Horst (2).

Edwards and Horst assumed that social desirability operated in such a way that the weights assigned to the items by subjects would be an increasing function of the social desirability scale values of the items. This assumption was based upon research by Edwards (1). He found that the probability of endorsement of an item in a personality inventory was an increasing linear function of the social desirability scale values of the items. No evidence was presented, however, to indicate that responses of subjects to items in Q sorts would be influenced in the same manner by differences in social desirability values of the items as was found in the case of responses to a personality inventory. The present study was undertaken to provide evidence upon this point.

Subjects consisted of a group of 50 men and a group of 50 women students at the University of Washington and at Oregon State College. Q sorts for the Oregon students were made available by Matthew J. Trippe. The procedure described by Stephenson (3, pp. 59–61) was followed in obtaining the Q sorts. Each subject gave a description of himself in terms of the 135 personality items used in the earlier study by Edwards. Eleven categories, ranging from most characteristic to least characteristic, were used by the subjects in making their self-descriptions. The frequencies for each category were fixed to resemble a somewhat flattened normal distribution. The number of items to be placed in the successive categories was 5, 7, 8, 14, 20, 27, 20, 14, 8, 7, and 5. The weights assigned to the categories consisted of the integers 1 to 11, with the 11 weight being assigned to the most descriptive category.

For each group of 50 subjects the mean weight assigned to each of the 135 items was found. These mean weights were then correlated with the social desirability scale values as determined previously by Edwards. The product-moment correlation for the male group was .84 and for the female group it was .87. If these results are typical of those to be expected in Q studies involving self descriptions in terms of personality items differing in social desirability values, then one might predict fairly successfully, on the average, the Q sorts of subjects, provided only that we have available the social desirability scale values of the items used in obtaining the Q sorts.

Received April 5, 1955.

References

1. Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
2. Edwards, A. L., & Horst, P. Social desirability as a variable in Q technique studies. *Educ. psychol. Measmt*, 1953, **13**, 620–625.
3. Stephenson, W. *The study of behavior.* Chicago: Univer. of Chicago Press, 1935.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

The Reliability of Adjustment Ratings and the Length of Case Histories

A. W. Bendig
University of Pittsburgh

Several methodological questions in the use of rating scales to quantify clinical judgments have been investigated in recent research (3, 4). Bendig and Sprague (4) had ten short clinical case histories, culled from psychology textbooks, rated for adjustment level by four groups of *S*s: introductory psychology students, students of abnormal psychology, graduate students in psychology, and post-Ph.D. clinical psychologists. The first two groups were subdivided into groups using rating scales differing in length: 3, 5, 7, or 9 categories. A progressive increase in rater reliability (average interjudge correlation) with increasing educational background was found. Also noted was a strong positive relation between scale length and rater reliability for the introductory psychology *S*s, but this relationship was not as evident for the abnormal psychology *S*s. In the second study, Bendig (3) used the longer case histories abstracted by Cummings (5) from clinical files. Two groups of *S*s, introductory educational psychology students and graduate students in psychology, rated the 10 cases for adjustment level using a 7-category scale. In addition, each *S* indicated which case history section was most important in determining his rating of the case. The groups were compared on the magnitude of several types of rating errors shown and on their use of various case sections.

The most obvious difference in the results of the two studies (3, 4) was in the magnitudes of rater reliability reported. Bendig and Sprague reported (4, p. 208) that the average intercorrelation among the introductory psychology *S*s using a 7-category scale was .47 ($N = 21$) while the reliability for the psychology graduate students ($N = 16$) was .59. The comparable reliabilities for the two similar groups in the second report (3) were .82 and .84. Changing the rated stimuli (case histories) not only increased the size of the rater reliabilities, but also tended to reduce the difference in reliability of these educationally distinct groups of judges. One hypothesis to explain these differences is that longer case histories (reporting more of the clients' behavior) would be rated more reliably, since the first group of cases were quite brief (approximately 100 words each) while the second set were much longer (one to two pages of single-spaced typing each). A second hypothesis suggests that the second set of cases, drawn from clinic files, covers a wider range of adjustment, or is more heterogeneous on the rated continuum (adjustment level) than is the first set, which was abstracted from textbooks. It has been shown (1, 2) that more heterogeneous stimuli are rated more reliably by introductory psychology *S*s.

The present research was designed to test this first hypothesis. The case histories in the second set were systematically shortened until they approximated the length of the first set. Both the "long" and "short" versions of the same cases were rated by different groups of introductory psychology *S*s which were further subdivided into groups differing in the length of the rating scale used. This latter step was designed to test for the appearance of the positive relation between scale length and rater reliability previously reported (4).

Procedure

Stimuli. Two sets of ten case histories each were used as stimuli to be rated. The "long" set consisted of the case histories originally compiled by Cummings (5) and later used by Newton (9) and Bendig (3). Each case had been abstracted from the complete clinic file, excluding psychometric information, with the *S*'s behavior reported under seven areas (3). The cases varied in length from one to two pages of single-spaced typing. In a previous study (3) both undergraduate psychology and graduate psychology students had rated these cases for adjustment level and had also indicated which of the seven case sections was most important in determining the adjustment rating they gave the case. The case sections checked by one-third or more of the raters in either group as being important have been reported (3). The "short" set of case histories used in the present study consisted of only the sections judged by either of the two previous groups as being important. The "unimportant" sections for each case were omitted in typing the cases for presentation to the raters. Each case in the "short" set was about one-half page of single-spaced typing, or approximately one-half to one-third its length in the long set. The cases of either the long or short set were mimeographed as a booklet with a face-sheet of rater instructions requesting the judge to read and rate each case for adjustment level, using a rating scale provided on a separate card.

Scales. Three different rating scales, differing in the number of categories on the scale (scale length), were mimeographed on cards with spaces for recording the judge's name, sex, age, and the adjustment rating given each case. The scales had either five, seven, or nine categories with the lowest category given a numerical weight of one and the highest category a weight of five, seven, or nine depending on scale length. Intermediate categories were given appropriate unit weights. The lowest category was verbally anchored by the descriptive phrase "Slightly maladjusted," the center category by the phrase "Moderately maladjusted," and the highest category by the phrase "Extremely maladjusted." The intermediate categories were left unanchored. These scales, both in length and in type of verbal anchoring, are identical with those used in previous research (3, 4).

Subjects. The judges were 120 undergraduate students enrolled in three sections of a course in introductory psychology. The case history booklets and scales were randomly distributed to the *S*s approximately midway in the semester with 20 *S*s getting each of the combination of case history length and scale length. The *S*s rated the cases outside of class and returned the booklets and scales at a subsequent class meeting. The usual request for noncollaboration among the *S*s was made in distributing the booklets and in the face-sheet of rater instructions.

Results

The case histories by *S*s matrix of ratings for each of the six subgroups (two case lengths by three scale lengths) were analyzed as a two-criterion analysis of variance, yielding mean squares for Cases ($df = 9$), Raters ($df = 19$), and Error ($df = 171$). The results of these six analyses can be found in Table 1.[1] The Cases mean squares were significant at the .01 level in each subgroup, indicating that the *S*s significantly discriminated between the case histories in their judgments of adjustment level. The Raters mean square for the subgroup using the 7-category scale to rate the long case histories was significant at the .01 level, while the similar mean square for the group rating the short cases was significant only at the .05 level. Neither of the Raters mean squares for the subgroups using 9-category rating scales were significant at the .05 level. Differences exist among raters in the average rating they assign all ten cases, but these rater differences, which have been previously termed "rater bias" (1, 2, 3, 4), were reduced with the longer scales.

The pairs of rater subgroups using similar length scales, but differing in the case length rated, were combined and three new analyses of variance, similar to the model described by Edwards (6, pp. 284–296), were computed. To test the feasibility of combining the pairs of subgroups, the ratios of the Error mean squares from the two groups were obtained and found not to be significant at the .10 level. The results of these analyses can be found in Table 2[1] with the tests of the homogeneity of the error variances given at the bottom of the table. The Cases mean squares were significant at the .01 level in all three analyses. The Between Groups mean square was significant at the .05 level for the subgroups using the 5-category scale with the subgroup rating the short case histories as showing more maladjusted behavior. However, the Groups mean square for the 7- and 9-category scales did not approach a significant level of confidence. The Cases by Groups interaction mean square was significant at the .01 level for the subgroups using the 7-category scale, but again the similar interaction terms for the subgroups using 5- and 9-category scales were not significant at acceptable levels.

---

[1] Tables 1, 2, 3, 5, and 7 have been deposited with the American Documentation Institute. Order Document No. 4674 from the ADI Auxiliary Publication

It has been previously noted (3) that Cummings (5) had six psychiatrists rank these ten case histories for adjustment level and derived scale values for the cases from the pooled ranks. The ratings of each of the 120 *Ss* in the present study were correlated with these scale values to give a measure of the "psychiatric reliability" of each *S*. The 20 product-moment correlations within each subgroup were transformed into normally distributed variate by the usual r-to-z technique and tested for intragroup heterogeneity (6, p. 135). None of the resulting six chi-square values was significant at the .10 level. The subgroup variances were then tested for intergroup heterogeneity by Bartlett's test (6, pp. 195–197) and the chi-square value (9.11) was not significant at the .10 level ($df = 5$). Finally, the z values were subjected to an analysis of variance and the results are reported in Table 3.[1] The variables of case length (long vs. short) and scale length (5-, 7-, or 9-categories) did not significantly affect the measures of psychiatric reliability and the interaction of these two variables was similarly insignificant.

Table 4

Summary of Average Individual Rater Reliability and Bias Coefficients for Each Group

Case history length	Length of rating scale	Psychiatric reliability	Rater reliability	Rater bias
Short	5	.80**	.64**	.52**
	7	.86**	.74**	.44*
	9	.85**	.78**	.34
Long	5	.83**	.73**	.62**
	7	.83**	.76**	.73**
	9	.84**	.73**	.28
Both	5	.82**	.68**	.60**
	7	.85**	.75**	.63**
	9	.85**	.75**	.29

* Significant at the .05 level.
** Significant at the .01 level.

From the data used for Tables 1, 2, and 3, average measures of psychiatric reliability (average correlation with psychiatric judgments), rater reliability (average correlation among the raters within a subgroup), and rater bias (variability among *Ss* in the average rating each *S* assigns all ten cases) were computed for each group. These measures are described in more detail elsewhere (3, 4). The summary values are given in Table 4. Both psychiatric reliability and rater reliability increase slightly (but nonsignificantly) from the 5- to 7-category scales, but no difference in these two measures is apparent between the 7- and 9-category scales. Rater bias decreased with longer scales. There appears to be no consistent difference between the long and short case histories on any of these three measures.

Mean ratings of the single cases were computed for each of the six subgroups and the case means were correlated with the psychiatric scale values referred to above. The intercorrelations of these mean ratings between the rater subgroups were also computed and these product-moment correlations are given above the principal diagonal in Table 5.[1] This matrix of intercorrelations between the rater and psychiatric subgroups was subjected to a centroid factor analysis. The analysis was replicated twice to stabilize the communality estimates and two factors were extracted. The residual correlations between groups can be found below the principal diagonal in Table 5 and the unrotated factor loadings and communalities for the seven

---

Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

groups are given in Table 6. The first general factor accounts for the great majority of covariation among the subgroups and appears to be a general "agreement" factor present in the mean case ratings of the rater subgroups. Factor II is bipolar with the psychiatrists and student raters using the 7-category scale with the short case histories on the negative end of the factor continuum and the three rater groups rating the long cases at the positive end. The groups using the 5- and 9-category scales to rate the short cases fall midway between these extremes. Apparently the *S*s rating the short cases were slightly more similar to the psychiatrists in their judgments than were the *S*s rating the long case histories.

In Table 7[1] are given the mean ratings given each of the ten cases by the six subgroups and *t* tests of the significance of the differences between mean ratings given the long and short cases by subgroups using the same length of scale. Standard errors for the *t* tests were derived from the error mean squares in Table 2. The probabilities from the three *t* tests for each case were combined by the chi-square technique of Fisher (7, 8). Three of the combined chi squares were significant at the .01 level (cases 7, 8, and 10) while two were significant at the .05 level (cases 3 and 5). In one case (case 10) raters receiving short case histories rated the case as less well adjusted than those rating long cases, while the other four cases were rated as being more well adjusted by the short case history subgroups.

Table 6

Factor Loadings (Decimal Points Omitted) for Rater Groups Derived from Intercorrelations of Mean Case History Ratings of Groups

Rater group	Case length	Length of scale	Unrotated factors		
			I	II	h^2
1	Short	5-Categories	988	015	976
2	Short	7-Categories	971	−128	959
3	Short	9-Categories	988	013	976
4	Long	5-Categories	982	103	975
5	Long	7-Categories	991	111	994
6	Long	9-Categories	991	087	990
7	Psychiatrists		948	−138	918

Discussion

The question of whether the elimination from case histories of material previously judged to be unimportant will adversely affect the reliability of judgments of adjustment derived from the case histories appears to be answered negatively. No significant differences were evident between the *S*s rating the "long" and "short" cases either in average intercorrelation among the *S*s (rater reliability) or their correlation with psychiatric judgments (psychiatric reliability). In fact, the factor analysis of the subgroup intercorrelations of mean case ratings suggested that the *S*s rating the short cases were somewhat more similar in their judgments to the psychiatrists than were the *S*s rating the long case histories. The elimination of the irrelevant material from the cases may have eliminated distracting information that the psychiatrists consciously excluded from consideration in forming their judgments, but that the more naive *S*s, reading the long cases, allowed to bias their ratings.

The difference in rater reliability found in one study that used short case histories (4) and a second study in which longer cases were used (3) apparently cannot be explained on the basis of case length differences. The length of the cases in the present research is comparable to the first study (4), but the magnitude of reliability is comparable only to those reported in the second study (3). Apparently the lower reliabilities reported by Bendig and Sprague (4) were due to the selection of the cases from psychological textbooks. These cases may have been much more homogeneous in adjustment level than are the present case histories and increased homogeneity of the rated stimuli has been shown to reduce rater reliability (1, 2).

The previously reported positive relationship between rating scale length and rater reliability (4) was not strongly evident in the present results. A small, nonsignificant increase in reliability from the 5- to 7-category scale was noted, but no difference was found between the 7- and 9-category scales.

This discrepancy between the two studies may be attributed either to omnipresent chance, or to the differences in the absolute magnitudes of rater reliability found. The present *S*s were about as reliable as judges can be expected to be, while the *S*s in the previous research showed relatively low reliability. The most tenable hypothesis is that there is a positive relationship between scale length and rater reliability, but that the strength of this relation decreases as the absolute magnitude of rater reliability increases. This hypothesis would also explain why a strong relationship was found for introductory psychology *S*s, while this relation was less apparent for more educationally advanced *S*s (4), since the less naive raters also displayed greater rater reliability.

The decrease in rater bias with longer scales contrasts markedly with the opposite result noted by Bendig and Sprague (4). No ready explanation of this discrepancy, other than the obvious chance or case differences between the two researches, appears plausible. Further evidence is needed on this point.

Summary

Introductory psychology *S*s ($N = 120$) rated for adjustment level long and shortened versions of ten case histories using scales with either 5, 7, or 9 categories. No differences were found between the long and short cases in measures of rater reliability, but a factor analysis of the intercorrelations of mean case ratings among the subgroups, and with the case ratings of psychiatrists, indicated that the ratings of the short cases were slightly more similar to the psychiatric judgments than were the judgments of the longer case histories. Rater reliability increased slightly, but nonsignificantly, between the 5- and 7-category scales, while no difference in reliability between the 7- and 9-category scales was evident. Rater bias tended to decrease with the longer scales.

Received April 5, 1955.

References

1. Bendig, A. W. Reliability of short rating scales and the heterogeneity of the rated stimuli. *J. appl. Psychol.*, 1954, **38**, 167–170.
2. Bendig, A. W. Rater reliability and the heterogeneity of the scale anchors. *J. appl. Psychol.*, 1955, **39**, 37–39.
3. Bendig, A. W. Rater experience and case history judgments of adjustment. *J. clin. Psychol.*, 1955, **11**, 127–132.
4. Bendig, A. W., & Sprague, Janine. Rater experience and the reliability of case history ratings of adjustment. *J. consult. Psychol.*, 1954, **18**, 207–211.
5. Cummings, S. T. The clinician as judge: the reliability and validity of judgments of adjustment inferred from Rorschach single-card performance. *J. consult. Psychol.*, 1954, **18**, 243–247.
6. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
7. Fisher, R. A. *Statistical methods for research workers.* (10th Ed.) New York: Hafner, 1948.
8. Fisher, R. A. Combining independent tests of significance. *Amer. Statistician,* 1948, **2**, 30.
9. Newton, R. L. The clinician as judge: the reliability and validity of judgments based on total Rorschachs and as applied to evaluations of clinical case material. *J. consult. Psychol.*, 1954, **18**, 248–250.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Some Evidence on the Validity of the Sarason Test Anxiety Scale

Barclay Martin and Bruce McGowan
University of Wisconsin [1]

Sarason (1) has developed a scale designed to measure how anxious a person feels in testing situations. This paper reports further evidence bearing on the validity of this test.

Two extreme groups of 18 *S*s each were selected from a class of 340 students, on the basis of a slightly modified Test Anxiety Scale, and then seen twice in individual sessions. Both sessions consisted of a 4-minute rest period followed by a 5-minute discussion of their feelings and attitudes about test-taking. Readings of palmar skin conductance were taken at regular intervals. The first session occurred early in the semester when no course examinations were being given, and the second session occurred within the hour preceding a course examination. Three things stand out in the results: (*a*) with the exception of the Session 1 rest period, the high-anxiety group had a significantly higher skin conductance for all periods than the low-anxiety group ($p < .05$); (*b*) it is also apparent that both groups tended to be relatively less apprehensive in the second session as indicated by decreases in skin conductance for both rest and discussion periods, although these decreases tended to be of only borderline significance; (*c*) there were no significant differences in relative change between the high- and low-anxiety groups from Session 1 to Session 2, and also no significant differences in relative change from rest to discussion period in either session.

These results suggest that Sarason's scale reflects partly a general anxiety factor which accounts for the general differences in conductance level between the groups. It can further be concluded that apprehension about the experimental session itself overshadowed concern about the course examination for both groups, as shown by the decrease in conductance from first to second session, presumably as a result of familiarity with the situation and knowing what was likely to happen in the session. These results raise the question of how specific to testing situations is the anxiety that the Sarason Scale measures, since there was no differential tendency for high as compared with low test anxiety *S*s to be more anxious right before a test than otherwise. It may be that most of the differences found in performance between high and low groups are due to a general anxiety factor rather than to specific test anxiety.

Brief Report
Received July 14, 1955.

Reference

1. Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 166–173.

---

[1] An extended report of this study may be obtained without charge from Barclay Martin, Psychology Dept., Univer. of Wisconsin, 600 N. Park St., Madison 6, Wis., or for a fee from the American Documentation Institute. Order Document No. 4641 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Movement as a "Rhetorical Embellishment" of Human Percepts

Leon H. Levy
Indiana University

Although the human movement response, *M*, has played a major role in Rorschach test interpretation—Beck describes it as opening "to investigation a sector of personality that, effective as it is in determining the individual's course, has been equally elusive of efforts to study it objectively" (2, p. 22)—few attempts have been made to test Rorschach's assumption regarding its nature. Thus most of the interpretive hypotheses based on amount of *M* in a Rorschach protocol are still predicated upon the assumption that *M* represents a projection by the subject of certain kinesthetic experiences into the blot (2, 3, 8, 9).

There is no dearth of speculation concerning how kinesthesis may be stimulated by the inkblot, or other static visual stimuli, for that matter (1, 7, 12). But as yet no empirical data confirm the occurrence of the conditions necessary for kinesthesia, viz., changes in the musculature of the individual. Indeed, as Rust (7) pointed out, Rorschach cautioned against scoring *M* in many instances even when the subject resorted to gestures to communicate his percept. It is therefore doubtful whether *M* and non-*M* responses will ever be distinguished on other than a verbal basis. But since Rorschach specifically differentiated between *M* responses resulting from "movement sensed in the figure" (6, p. 25) and non-*M* which are "only a rhetorical embellishment of the answer, a secondary association" (6, p. 25), it is clear that the scoring criteria for *M* must include some nonverbal measure. Since nonverbal indices did not prove expedient, Rorschach and those who have followed him have reasoned that one could sense movement only if one perceived the movement as occurring in a human figure, so that only human movement is scored *M*. The scoring is based on the concept of empathy, whereby it is assumed that a person cannot empathize with animals or inanimate objects, but only with other humans. Since the concept of empathy is itself not clearly defined, the principle does little to clarify theory or practice.

To those who prefer concepts somewhat better defined than *M* is, there appear to be two alternatives. The one is to discard the term entirely, while the other is to redefine it in terms which are more verifiable. In many instances the former alternative is to be preferred. In the case of *M* there is some evidence (4, 5, 10, 11), albeit not entirely unequivocal, of meaningful relationship to other psychological phenomena, so that the more conservative approach would appear justified. *M* may be a valuable correlate of certain psychological phenomena regardless of whether or not Rorschach's conception of its nature was correct or whether or not his interpretation is testable. If some more adequate conceptualization of *M* is possible, its interpretive value may be further enhanced.

The explanation offered in this paper is one with which Rorschach would undoubtedly disagree. It is proposed that *M* is best understood as a "rhetorical embellishment" or "secondary association" to use Rorschach's own terms; that it is an epiphenomenon of the tendency to perceive human figures in the stimulus material.

This view of *M* is predicated on the hypothesis that when a person is asked to describe a human figure or a stimulus taken as representing a human figure, there is a high probability that movement of one kind or another will be among the attributes cited. Describing any object is accomplished by listing its attributes. Since one of the most frequently reported attributes of the human organism is movement, the taking of some position, engaging in some kind of activity, etc., it should be expected that these attributes will be among those cited by the individual in his report of his perception of any visual stimulus he takes to represent the human figure.

Hypothesis

The frequency of occurrence of verbal responses classifiable as movement responses in subjects' reports of their perception of certain visual stimuli having no common cultural interpretation will be directly related to whether or not these stimuli are presented as representing human figures.

Procedure

The 112 subjects (*S*s) used in this study consisted of all those members of four discussion sections of an introductory psychology course who were present on the day of the experiment.

Ten stimulus figures were constructed by combining three lines: two one and one-half inch lines joined to form a 60-degree angle, and one two-inch line drawn so that one of its end points was the apex of the angle formed by the other two lines. The figures were varied by varying the angle of the longer line in relation to the vertical, and also by varying the size of the angle formed by the longer line and either of the two shorter lines. The figures were presented to *S*s in booklet form with one stimulus to a page.

The *S*s were divided into two groups of 56 each, each group consisting of two of the discussion sections. Group H received booklets with the following instructions on the front page:

> This is a test to see how imaginative and original you can be. In the booklet before you, you will find a series of ten figures composed of three lines in various arrangements. *Each of these figures is meant to represent a person.* You are to use your imagination and describe in one sentence under each figure what the figure might represent. Of course there are no right or wrong answers to this; the important thing is to give the most imaginative descriptions possible.

The *S*s in Group NH received identical instructions except that the italicized sentence was deleted. In each case these instructions were also repeated orally by the experimenter. All responses were scored for *M* using Klopfer's criteria (3).

Results

The mean number of *M* responses for Group H was 8.62 as compared with 1.12 for Group NH. Because of the nonnormal nature of the distribution, chi square was used in testing the hypothesis. Chi square was calculated for a 2 × 2 table comparing the number of cases in each group falling above and below the median of the total sample, 3.5. With Yates's correction for continuity, chi square was 82.312, significant beyond the .001 level.

Discussion

The results seem to support the interpretation that amount of *M* in a Rorschach test protocol is a function of the tendency of the individual to perceive human figures in ambiguous visual stimuli. In one sense, the finding comes as no surprise since *M* can only be scored for human percepts, but it causes one to question what more, if anything, can be said regarding the nature of *M*.

What is to become of the various interpretations which are dependent upon *M*? To the extent that interpretations are dependent only upon rational exposition and lack empirical justification, they may be discarded with impunity. To the extent that they have been found to be empirically valid, they pose new and interesting research problems centering around the question: What is the psychological significance of a tendency to perceive human figures in ambiguous visual stimuli? Since it appears from the findings that *M* does not account for all of the variance of *H*, further research might be directed toward studying instances where an individual gives a non-*M* human response.

The "scoring" of a Rorschach test protocol might best be construed as a problem in the analysis of verbal behavior. The problem faced in Rorschach test research is one of determining what the most useful units of analysis shall be and how they are related to other psychological phenomena. The findings suggest that unless *M* can be shown to be consistently related to other psychological events in a way which is independent of other scores, particularly the content score *H*, it might best be discarded. The result should be a simpler classification system of the same analytic power.

Summary

An experiment was performed to test the hypothesis that the tendency to give verbal responses scoreable as human movement, *M*, by Rorschach test scoring standards, is directly related to the extent to which the individual takes the visual stimulus to represent a human figure. The justification for scoring both *M* and *H* in Rorschach test analysis was questioned in the light of findings confirming the hypothesis.

Received March 14, 1955.

References

1. Arnheim, R. Perceptual and aesthetic aspects of the movement response. *J. Pers.*, 1951, **19**, 265–281.
2. Beck, S. J. *Rorschach's test.* Vol. II. New York: Grune & Stratton, 1946.
3. Klopfer, B., Ainsworth, Mary D., Klopfer, W. G., & Holt, R. R. *Developments in the Rorschach technique.* Vol. I. Yonkers, N. Y.: World Book Co., 1954.
4. Prados, M. Rorschach studies on artists-painters. I. Quantitative analysis. *Rorschach Res. Exch.*, 1944, **8**, 178–183.
5. Roe, Anne. Painting and personality. *Rorschach Res. Exch.*, 1946, **10**, 86–110.
6. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
7. Rust, R. M. Some correlates of the movement response. *J. Pers.*, 1948, **16**, 369–401.
8. Sarason, S. B. *The clinical interaction.* New York: Harper, 1954.
9. Schachtel, E. G. Projection and its relation to character attitudes and creativity in the kinesthetic response. *Psychiatry,* 1950, **13**, 69–100.
10. Singer, J. L., & Herman, J. Motion and fantasy correlates of Rorschach human movement responses. *J. consult. Psychol.*, 1954, **18**, 325–331.
11. Tucker, J. E. Rorschach movement responses and intelligence. *J. consult. Psychol.*, 1950, **14**, 283–286.
12. Werner, H. Motion and motion perception: A study of vicarious functioning. *J. Psychol.*, 1945, **19**, 317–327.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

The Measurement of Primary Mental Abilities by the Columbia Mental Maturity Scale[1]

Cyril R. Mill and Charles J. Turner
Richmond Public Schools, Richmond, Virginia

The Columbia Mental Maturity Scale (CMM), published in 1954, was designed to measure the mental ability of children in the MA range from 3 to 12 years. Since the only response required is to point, it was thought suitable for use with children afflicted with cerebral palsy or other handicaps involving motor or verbal functioning. Administration and scoring are simple and testing time runs about 15 minutes. It is a potentially useful test for a quick measure of intelligence in clinic and school use with normal children. This study investigated the validity of the CMM IQ's, factors of intelligence measured by the CMM, and age levels at which it functions best.

Seventy-five normal (nonhandicapped) children, equally divided by sex, from ages 7–0 to 11–11 years, were administered the CMM and the SRA Primary Mental Abilities tests. CMM IQ's correlated .63 with PMA Total IQ's and .62 with PMA Non-Reading IQ's. It was therefore concluded that the CMM did not provide a validity coefficient in this instance sufficiently high to warrant its use as an individual measure of intelligence.

Correlations between CMM and the five subtests of the PMA were as follows: with Reasoning, .61; with Verbal Meaning, .51; with Perception, .48; with Number, .41; with Space, .31. These *r*'s suggest that the CMM is most heavily weighted with a measure of reasoning and understanding of ideas. It contains a moderate measure of numerical ability and ability to perceive small details quickly and accurately, and there is a slight measure of space perception.

The mean CMM IQ's were higher than mean PMA Total IQ's at all age levels tested. The differences were not significant at the nine- and ten-year levels. The CMM IQ's were significantly higher for the eight year olds (5% level) and for the 7 and 11 year olds (1% level). The clinical impression was that an excessive number of numerical analogy problems was used, with the result that some children learned to solve the tasks on the analogy principle without regard to the content.

The *SD* of the CMM IQ's was twice as large as the *SD* of the PMA Total IQ's. Scores for boys were significantly more variable than for girls, possibly due to differences in social maturity of the sexes at these age levels.

This study indicates that the CMM measures reasoning and verbal meaning even though it is a nonverbal test. It is not sufficiently valid at present for clinical use, but with further refinement it could fill a real need for a quick IQ measure for both normal and afflicted children.

Brief Report.
Received August 23, 1955.

---

[1] An extended report of this study may be obtained without charge from Cyril R. Mill, 407 N. 12th Street, Richmond, Va., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4679 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Reactions to Rorschach Cards IV and VII as a Function of Parental Availability in Childhood

Ralph Hirschstein and Albert I. Rabin
Michigan State College

Rorschach workers have for some time been aware of the symbolic value and interpretational importance of individual inkblots. Particular attention has been paid to Cards IV and VII as symbolizing the father and mother figures, respectively. Despite this widespread practice, there is little explicitly stated in the literature, in support of this symbolic interpretation. Bochner and Halpern (1), speaking of Card IV state that "The heavy male figure may suggest the father or authority. . . ." In Card VII they see ". . . a feminine quality, frequently with maternal implications." More recently, some experimental evidence has been marshalled in support of the symbolic meaning of the aforementioned Rorschach cards (2, 3). This evidence is based upon direct questioning of subjects regarding the father and mother figures which might be represented by the inkblots. The results are based entirely on conscious association and critical reflection of the individuals tested.

The purpose of the present investigation was to make an indirect check on the symbolic value of Cards IV and VII of the Rorschach series. The main hypothesis was that individuals who are adolescent delinquents and in whose early background there were no significant mother or father figures would react more readily and more easily to these cards than would a similar group of delinquents who grew up in the standard family situation and who had the opportunity of acquiring conceptions of father and mother figures, disturbing as they may be.

Method and Subjects

The subjects of the present study were 40 male juvenile delinquents, divided into two groups of 20 each. The two groups were matched for age and intellectual level of functioning, the mean ages being 15.2 and 15.3 years, respectively, and the mean IQ's 97.6 and 96, respectively. All subjects had been placed by the courts at the Boys' Vocational School at Lansing, Michigan. None of the delinquents used in this investigation revealed any psychotic manifestations or organic involvement upon psychiatric examination. Group NF (Non-Family) consisted of twenty youngsters who had no opportunity of living with their natural parents and who had been moved around from one foster home to another since infancy. The F (Family) group consisted of twenty youngsters who had been living with their own families (parents) continuously until the time of their arrest. Causes precipitating arrest were similar for both groups. Most common reasons for arrest were in order of frequency: car theft, breaking and entering, and school truancy.

Rorschach records were obtained on all subjects under standard conditions of examination. Analysis of the data concentrated on the differences between the two groups with respect to Cards IV and VII. In accordance with our hypothesis, the expectations were that the "family" group would show greater difficulty in responding to Cards IV and VII than the "non-family" group. The underlying assumption is that part of the delinquents' difficulty lies in the intrafamily relationships which may be reflected in Cards IV and VII. This should not prevail for the "non-family" group who had no definite parental relationships.

Results

The two groups were compared with respect to form quality ($F+\%$), productivity (R),

Table 1

Significance of the Difference of Percentage of the Total Number of Responses to Cards IV and VII

Family group	Non-family group	t	Level of significance
16%	18.5%	1.68	10%

and "reaction time to first response" on Cards IV and VII. A comparison of the $F + \%$, for Cards IV and VII, yielded no significant differences. When the mean productivity for Cards IV and VII, combined, was compared for the two groups, a slight trend favoring the "family" group was noted. (See Table 1.)

A comparison of reaction times to the individual Cards IV and VII indicated differences in the expected direction. The mean reaction times for both groups, for all ten Rorschach cards, is given in Table 2. The only differences that were found to be statistically significant are those obtained on Cards IV and VII. The "family" group reacted more slowly to these cards than did the "non-family" group.

Table 2

Small Sample Tests of Significance of Difference of Means of Reaction Times to All Ten Rorschach Cards for the Family and Non-Family Groups

Cards	Family group Mean R.T.	Non-family group Mean R.T.	t	Level of significance
1	20.55	15.20	.56	insignificant
2	17.20	13.50	.59	insignificant
3	13.10	6.80	.95	insignificant
4	31.85	15.90	2.01	2%–3%
5	13.70	9.45	.89	insignificant
6	32.85	30.05	.31	insignificant
7	22.45	12.85	1.91	5%
8	19.15	17.85	.11	insignificant
9	26.90	21.95	.07	insignificant
10	16.80	13.55	.95	insignificant

It appears that the youngsters with ordinary family background were "shocked" or blocked when responding to Cards IV and VII; the youngsters with no parental background did not show this blocking. It would seem that one can differentiate youngsters lacking a stable parental figure from those possessing one, on the basis of reaction time to Cards IV and VII.

Summary

Two groups of male delinquent subjects, matched for age and intelligence, but differing with respect to parental availability in early childhood, were compared on the basis of their reactions to Rorschach Cards IV and VII. The youngsters who grew up with their natural family reacted significantly more slowly to these cards than did the group who had no real mother or father figure to identify with. Furthermore, a slight trend in the direction of restricted productivity to these cards, was noted in youngsters who had grown up with their natural parents. These results are offered as further evidence in support of the hypothesis that Cards IV and VII symbolize parental figures.

Received March 22, 1955.

References

1. Bochner, Ruth, & Halpern, Florence. *The clinical application of the Rorschach test.* New York: Grune & Stratton, 1945.
2. Meer, B., & Singer, J. L. A note on the "father" and "mother" cards in the Rorschach inkblots. *J. consult. Psychol.,* 1950, 14, 482–484.
3. Rosen, E. Symbolic meanings in the Rorschach cards: a statistical study. *J. clin. Psychol.,* 1951, 7, 239–244.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Rorschach Content and Hostile Behavior[1]

Martin R. Gluck
Mental Hygiene Consultation Service, Fort Lewis, Washington

A question of major concern in the use of projective techniques, and one which has provoked arguments both pro and con, is that of the relationship between the content of such protocols and the testee's overt behavior. Piotrowski (8), Wittenborn (11), and Rosenzweig (9), for example, either state or imply that direct prediction from projective records to overt behavior is possible. Klopfer's most recent Rorschach manual (7) discusses these problems of prediction and content analysis much more fully than did his older book (6), while Beck (1), too, in one of his recent reviews of the Rorschach technique, stresses the need for further research into the meaning and importance of content. Lastly the author's work-a-day experience suggests that users of projective tests, both psychologists and psychiatrists alike, frequently tend to make such predictions and to act in accordance with them.

Hostility, especially its direction and strength, is one of the major dynamic forces which the projective tester may be called upon to assess through the medium of the Rorschach technique. The present study is therefore focused on the question: Can hostile (aggressive) behavior be predicted from the amount of hostility (aggression) found in the content of Rorschach protocols?

Prior research of this "predictive" kind has used criteria which have differed widely, not only from this paper but from one another. In an early study Young and Higgenbotham (12) searched for agreement in major psychological areas between Rorschach protocols and a large mass of behavioral data (e.g., case histories, counselors' records, etc.), concluding that, although single determinants and simple ratios were not very useful, the Rorschach does reflect " . . . potential forces and tensions within the individual . . . the resolution of [which] appears in behavior relative to specific environmental stimuli" (12, p. 93).

Elizur (3), with a scoring system for evaluating both anxiety and hostility as expressed in Rorschach content, compared the amount of projected hostility with measures of other kinds of hostility based on a questionnaire, a self-rating scale, and the judgments of observers watching the subjects' behavior in an interview situation. He reported significant correlations between all of the variables used and, in addition, demonstrated that the use of the content scores could distinguish between the records of normal subjects and neurotic ones. Following Elizur's lead, Walker (10) constructed a more refined scale and reported significant correlations between Rorschach content scores and therapists' evaluations of hostility in their patients.

Problem and Definitions

Because of the divergence of the criteria used in the studies reviewed and also because of the differing kinds of scale systems used to evaluate Rorschach content, it seemed useful to carry out a further study in which the criterion situation (and the "predictor" scale) were tailored to the specific problem being investigated.

---

[1] Based on a dissertation submitted to the Graduate School, University of Pittsburgh in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author acknowledges the help and guidance of his dissertation committee, and especially thanks Doctors A. David Lazovik and A. W. Bendig. Thanks are also due to the members of the Psychology Section, Neuropsychiatric Service, Fitzsimons Army Hospital, and to Mrs. Elizabeth J. Scheide for rescoring the Rorschach records. The statements and conclusions of this study are the author's and do not necessarily reflect the opinion or policy of the Army Medical Service.

Therefore, the following three hypotheses were formulated.

1. The more overt hostility a subject expresses in his projective records, the more he will react with hostile behavior when placed in a situation which is provocative of such behavior.

2. The more covert hostility a subject expresses in his projective records, the less he will react with hostile behavior when placed in a situation which is provocative of such behavior.

3. Subjects who have a large amount of hostility of both kinds in their records will react with more hostile behavior than subjects whose total amount of hostility (in their records) is relatively less.

"Behavior" appears here in the sense of gross actions or activities which are readily observable by another individual. "Hostility" is also used in the broadest fashion; that is, it is meant to connote feelings and/or actions carrying the idea of such words as anger, hatred, resentment, irritation, and exasperation.

Method

Subjects in this research were thirty patients of the Neuropsychiatric Service, Fitzsimons Army Hospital, who were given a standard test battery as an integral part of their individual psychological evaluations; the Rorschach was administered to each subject individually and in each examiner's usual manner. The examiners were all psychology interns in training at the hospital. The author administered half of the total number, while the remaining protocols were obtained by the other four interns.

After the psychological testing was completed, the subjects were put into a stress situation designed and administered in such a way as to be extremely provocative of hostile behavior. There were nine tasks in the stress situation, wherein the subject's behavior and reactions to each were rated at the conclusion of the task by each of three observers. The final criterion score (of hostile behavior) was a composite derived from the observers' pooled judgments of each subject on three of the tasks. Statistical tests indicated that the observers' judgments of hostility were reliable and discriminating.[2]

The Rorschach (Hostility) Content Rating Scales

Although there was some recognition of the difference between openly expressed and indirectly or covertly implied hostility in both Elizur's and Walker's work, these writers did not directly attack the problem as such. Consequently, it was decided to attempt an appraisal of these seemingly (qualitatively) different expressions of hostility. Using examples and definitions gleaned from the aforementioned two studies and from De Vos's (2) paper, scales were constructed, for both Overt and Covert expressions, with values ranging from 0 to 2. Before the scales were useable, it was found necessary to redefine the points on the scales jointly with the other rater. Sample responses from the sources mentioned above were employed as examples for the various scale points. This joint defining of the scale points should be borne in mind when evaluating the reliability figures presented in Table 1.

"Overt" responses are those in which hostile, aggressive feeling is plainly conveyed by the content, thereby evoking little or no interpretation by the clinician before he labels these responses as hostile ones. Such responses have content which carry a commonly accepted (i.e., "everyday") hostile or aggressive connotation.

"Covert" responses are those in which the hostility is perceived by means of symbolic interpretation, and which have content that does not usually convey a commonly accepted connotation of hostility. Examples of each point on both of the scales used are presented below.

For the Overt Scale responses such as "Bulls charging each other," "Two fellows in a duel—really fighting—you can see the sabres," and "Animals shot" received a rating of 2 (Most Intense).

A rating of 1 (Less Intensive) was given for such responses as "Spears used by Indians," "A quarrelsome person," "Pistol," "Arrow," "Gun," and 'Knife."

Similarly, a 2 rating on the Covert Scale was given for responses like "Drunk who fell on his

---

[2] A somewhat more detailed discussion of the criterion measure has been previously reported (5).

Table 1

Reliabilities of Rorschach Scales
($N = 30$)

Rating	Overt	Covert	Total
Independent Ratings	.83	.67	.68
Combined Ratings	.91	.80	.81

back," "Birds dressed in men's clothing," and "Bat with a torn wing."

The Covert rating of 1 was applied to responses similar to "A volcano," "Teeth," and "Torn leaves."

The foregoing are not all of the examples used to define the scale points. In addition to further examples, the scales in their final form included statements of definitions of each of the rating points.

Each rater scored the full Rorschach protocols, scoring each response on whichever scale was appropriate. By adding the scores given on each of the scales, the subjects received three content hostility scores: an Overt score, a Covert score and a Total Score (Overt plus Covert scores). The reliability coefficients presented in Table 1 were computed from these three sums. The final Rorschach Hostility Scores were obtained by summing each of the three scores arrived at by each of the two raters, permitting the use of the Spearman-Brown formula (4, p. 390) and resulting in heightened reliabilities for the combined ratings, and then dividing the combined score for each scale by the number of responses in the record. In recognition of the fact that the significance of a response or series of responses is frequently judged in terms of the length of the record, it was felt that the use of such a ratio score was desirable.

The correlation between the Overt and Covert Rorschach Hostility scores was .33, a value which approaches significance at the .05 level. Apparently, the two scales used were not independently tapping qualitative differences.

Results and Discussion

The results of the comparison between Rorschach Hostility Scores and the Criterion Score (composite rating of behavioral expressions of hostility in the stress situation) show no significant correlations, the correlation coefficients between the Criterion Score and the Overt, Covert, and Total scores being .07, .18, and .16.

Contrary to the previous studies already cited, none of the three hypotheses of this study receive support from these data. The first consideration which comes to mind in comparing these findings with those of previous investigators is the difference in the criteria measures used. Walker worked with material derived from therapy, while Elizur used other test data and judgments from a relatively benign situation. The question then arises which, if any, of the methods is most appropriate or useful from the point of view of their relations to the patients' environmental situations. Walker's approach has merit from the standpoint of a therapist asking for help in planning the *therapeutic* situation, but neither his nor Elizur's criteria seem to bear directly on broader life situations, e.g., those incidents where the patient is faced by the threat of a strong frustration and/or an authority figure.

It was felt that in the research reported here the presence of a hostile and hostility-provoking authority figure, such as was the administrator of the stress situation, might well make the situation similar to those in which patients frequently have difficulty adjusting. However, this strong authoritarian examiner may well have provoked the anxiety necessary to inhibit the hostile action at the same time that he was provoking hostile feelings within the subjects.

This leads directly to a second consideration, i.e., the overlap of the functioning of the Overt and Covert Scales. As was noted earlier, the Rorschach content scales used here are not nearly as independent as one might wish. Both the author and the second scorer were impressed with the possibility that the sample responses might reflect anxiety feelings in addition to those of hostility in a kind of intermingling that has been noted by De Vos (2) and Elizur (3). Such a mutual relationship would have much significance for the type of prediction attempted in the present study, for without considering the role of anxiety as an inhibiting (as well as provoking) agent it appears that adequate prediction will not occur.

Conclusions

From the foregoing data it is concluded that simple assessment of the *amount* of hostility contained in the content of a Rorschach protocol does not provide an accurate index to the patient's proclivity or ability to behave in a hostile manner in what seems to be a hostility-provoking situation.

Other pertinent factors related to the amount of anxiety present, effective controls available (e.g., Form Level, quality of *M*, *FC* to *CF*, etc.), and direction or target of hostility will have to be more precisely described before we will have accurate prediction of the type sought for here. In this respect the writer differs with Korner's view (in Klopfer **7**, pp. 490–491) that projective techniques are most valuably used in ". . . fantasy exploration . . ." and concomittantly that prediction of behavior is not necessarily desirable. Instead, the orientation expressed in Ainsworth's introduction to the same chapter (**7**, Ch. 14) is more congenial to that of this paper inasmuch as it is felt there is usefulness in research on the predictive abilities of interpretive hypotheses. Although the results presented here certainly evoke reservations about such an application of content rating scales, the technique used in establishing a criterion measure does seem to afford a workable approach to the problem of receiving reliable judgments of "emotional behavior."

Summary

1. Hypotheses relating hostile Rorschach content to hostility expressed through behavior were stated.

2. Subjects were given a Rorschach examination and then put into a hostility-provoking stress situation, where the amount of hostility displayed behaviorally was rated by three judges.

3. The content of the Rorschach records was rated on scales which were designed to assess amount and type of hostile content.

4. Correlations between scores from the Rorschach content scales and the subjects' behavior were found to be nonsignificant.

5. Pertinent points were considered in relation to these findings and suggestions for the direction of future research offered.

6. It was concluded that the hypotheses offered were not supported by the data resulting from the study.

Received May 2, 1955.

References

1. Beck, S. J. The Rorschach test: A multi-dimensional test of personality. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice Hall, 1951. Pp. 101–122.
2. De Vos, G. A quantitative approach to affective symbolism in Rorschach responses. *J. proj. Tech.*, 1952, **16**, 133–150.
3. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, **13**, 247–284.
4. Garrett, H. E. *Statistics in psychology and education*. (3rd Ed.) New York: Longmans, Green, 1947.
5. Gluck, M. R. The relationship between hostility in the TAT and behavioral hostility. *J. proj. Tech.*, 1955, **19**, 21–26.
6. Klopfer, B., & Kelley, D. M. *The Rorschach technique*. Yonkers, N. Y.: World Book Co., 1946.
7. Klopfer, B., Ainsworth, Mary D., Klopfer, W. G., & Holt, R. R. *Developments in the Rorschach technique*. Yonkers, N. Y.: World Book Co., 1954.
8. Piotrowski, Z. A. A new evaluation of the Thematic Apperception Test. *Psychoanal. Rev.*, 1950, **37**, 101–127.
9. Rosenzweig, S. Levels of behavior in psychodiagnosis with special references to the Picture-Frustration Study. *Amer. J. Orthopsychiat.*, 1950, **20**, 63–72.
10. Walker, R. G. A comparison of clinical manifestations of hostility with Rorschach and MAPS test performances. *J. proj. Tech.*, 1951, **15**, 444–460.
11. Wittenborn, J. R. The implications of certain assumptions involved in the use of the TAT. *J. consult. Psychol.*, 1950, **14**, 216–225.
12. Young, R. A., & Higginbotham, S. A. Behavior checks on the Rorschach method. *Amer. J. Orthopsychiat.*, 1942, **12**, 87–94.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

A Statistical Comparison of the WISC and Wechsler-Bellevue, Form I[1]

John R. Price and Gareth D. Thorne[2]
University of Denver

Few studies making a comparison between the WISC and Wechsler-Bellevue Form I in an experimental situation have been reported. The need for such studies seems apparent insofar as both instruments are being used currently on the assumption that they are equivalent in the overlapping age range, 10 to 15 years. In clinical situations the choice of which of these two tests to use with children in this age range has been left to the judgment of the *E* who has little recourse to comparative studies to aid in making this decision. Because of the widespread use of these two scales in clinics and school systems, it is felt that the assumption of equivalence warrants further investigation. The hypothesis as set forth in this investigation is that the WISC and the WB Form I are not equivalent instruments with respect to the Verbal, Performance, and Full Scale scores within the aforementioned age range.

Graham (2) dealt with the problem of equivalence insofar as his sample of unsuccessful readers had been evaluated in part through the use of these two tests. This investigation, however, was primarily concerned with subtest analysis and did not statistically treat the Scale scores. In a study using mentally retarded children, Vanderhost *et al.* (4) used a design similar to the one employed in the present investigation and made comparisons of the means of each Scale score with every other Scale score of the tests used. In an investigation with normals, Delattre and Cole (1) were primarily interested in various comparisons of subtest scores but reported the obtained Scale scores in their article.

Method

It was considered advantageous to employ an experimental design that would permit the use of one group of *S*s for both tests. In order to prevent practice effect from either test influencing the results, the test to be administered first to each *S* was determined using a table of random numbers with the restriction that each test be administered first an equal number of times.

The criterion for selecting *S*s was solely chronological age. The lists of *S*s were compiled in such a manner that at the anticipated time of testing each *S* would be at least 11 years, 4½ months or 14 years, 4½ months of age respectively, and at most 11 years, 7½ months and 14 years, 7½ months in each age group. The sample was selected from local school systems and consisted of 40 white, American children at each age level. Each group consisted of 20 boys and 20 girls. These groups were divided so that both *E*s administered tests to 10 boys and 10 girls at each age level with each *E* administering both tests to each *S* in his group. Those *S*s who had taken either the WB Form I or the WISC previously were dropped from the group.

Both tests were administered to the *S* during one session with a standard rest period of 15 minutes between tests. In the actual testing all the subtests of both the WISC and WB were employed with the exception of the Mazes subtest on the WISC. Scoring of each test was done by the *E* who had administered the test. In the case of questionable responses on such subtests as Comprehension, Similarities, etc., a consultation was held between *E*s in order to arrive at a score that was satisfactory to both *E*s in terms of the criteria as set forth in the manuals.

---

[1] The writers wish to express their indebtedness to Drs. E. Ellis Graham, Joel E. Greene, and William L. Sawrey whose helpful advice and assistance made this study possible.

[2] Mr. Thorne is at present Assistant Director of Training at the Caswell Training School, Kinston, North Carolina.

Table 1

Means and Standard Deviations of Experimental Groups

Scale	WISC		WB	
	M	*SD*	M	*SD*
$11\frac{1}{2}$ Years*				
VS	102.98	13.36	94.90	14.25
PS	103.63	15.11	108.50	14.74
FS	103.65	14.25	102.15	14.87
$14\frac{1}{2}$ Years*				
VS	108.00	14.21	102.58	15.08
PS	108.25	9.68	110.23	10.84
FS	108.55	10.40	106.93	12.79

* The actual mean ages for these groups were 11 years, $6\frac{1}{2}$ months and 14 years, 7 months.

Results

At both the $11\frac{1}{2}$- and the $14\frac{1}{2}$-year age levels, higher mean IQ's were obtained on the VS and FS of the WISC than on the WB, whereas the mean IQ of the PS of the WB was higher than that of the WISC at both of these age levels, as shown in Table 1. The higher means on the VS and FS of the WISC and higher mean PS of the WB parallel the findings of Graham (2) and Vanderhost *et al.* (4), but do not agree with Delattre and Cole (1).

The initial comparisons involved testing for significance of differences between the experimental groups and the groups that were used by Wechsler to standardize both tests. Critical ratios between means of the experimental and standardization groups and the *F* ratios between variances for these groups are given in Table 2.

In order to test the hypothesis of lack of equivalence between two tests it is necessary to establish criteria that will define what is meant by equivalence. A technique used by Hollingsworth (3) for determining such criteria was employed. This method specifies two conditions that must be met in order to state that the two tests are equivalent. The first condition is that individuals should obtain essentially the same ranking on the VS, PS, and FS on one test as they obtained on the other allowing for chance error.

In determining whether or not the first condition has been met by the data, it is necessary to agree upon some arbitrary value of r for a specific criterion. In arriving at this arbitrary value for r it seemed that a relationship between the standard error of measurement of the WISC and the standard error of estimating WISC scores from WB scores would prove to be the most satisfactory. Since a perfect r between the two tests within each age group would be neither demanded nor obtainable, it was decided that "the desired r would be of such magnitude that the standard error of estimate would be not greater than 1.5 times the standard error of measurement" (3, p. 14) of the WISC. The following equation was set up to show this relationship:

$$\sigma_{\text{WISC}} \sqrt{1 - r^2_{\text{W.W}}} = 1.5\ \sigma_{\text{WISC}} \sqrt{1 - r_{\text{WISC reliability}}},$$

where $r_{\text{W.W}}$ is the desired r between the WISC and the WB. After solving this equation algebraically for $r_{\text{W.W}}$ the desired r was computed for all three scales at both age levels. To satisfy the first condition the desired r should lie between the obtained r and the upper 5% confidence limit for this obtained r. Table 3 indicates the results obtained in meeting this first condition.

The second condition necessary for equivalence is that each individual should obtain essentially the same IQ scores on one test as he obtained on the other, also allowing for chance error. In meeting this second condition, a technique devised by Wilks (7) was

Table 2

Comparisons Between Experimental and Standardization Groups

Scale	WISC		WB	
	Sig. of diff. between means	Sig. of diff. between variances	Sig. of diff. between means	Sig. of diff. between variances
$11\frac{1}{2}$ Years				
VS	1.21	1.11	1.72	1.35
PS	1.51	2.11**	2.54*	1.38
FS	1.27	2.21*	.44	1.12
$14\frac{1}{2}$ Years				
VS	3.28	1.34	.88	1.13
PS	4.97**	1.29	4.04**	1.70
FS	3.98**	3.88**	2.61**	1.31

* Significant beyond the 5% level.
** Significant beyond the 1% level.

Note.—In the Experimental Groups, $N = 40$ for both tests at both age levels. In the WISC Standardization Groups, $N = 200$ at both age levels; in the WB Standardization Groups, $N = 60$ at the $11\frac{1}{2}$-year age level, and $N = 70$ at the $14\frac{1}{2}$-year level. In testing the significance of differences between variances, $df = N - 1$.

Table 3

Comparisons Between WISC and WB of the Experimental Groups

Scale	Obtained r	Upper 5% confidence limit for obtained r	Desired r	L_{mvc}	L_{vc}	L_m
			11½ Years			
VS	.85	.92	.95†	.466**	.998	.467**
PS	.79	.88	.87	.797*	.999	.798**
FS	.89	.94	.94	.952	.998	.954
			14½ Years			
VS	.90	.95	.95	.592**	.996	.594**
PS	.41	.64	.88†	.957	.987	.970
FS	.78	.88	.93†	.918	.958	.958

* Significant beyond the 5% level.
** Significant beyond the 1% level.
† Failed to meet the first condition.

employed wherein two or more alternate forms of a test may be tested for equivalence. The initial procedure (L_{mvc}) involves testing the means, variances, and covariances simultaneously for equivalence. Provided the null hypothesis is rejected at the desired level of confidence, the formulation is then fragmented to determine whether the means, the variances, or the covariances are contributing to this lack of equality and if so to what degree. In the present investigation the test for variance-covariance (L_{vc}) reduces to simply a test between variances since there are only two groups present. However, the correlational term in the covariance has been taken into account in the first condition above. Wilks has demonstrated that the test for means (L_m) is equivalent to an analysis of variance design with the tests (WISC and WB) as columns and the individuals as rows. Consequently, the over-all test for equivalence (L_{mvc}) is a powerful and appropriate test to be used in the initial phase of the investigation. If there are no significant differences using this initial test, it is highly unlikely that the follow-up tests (L_{vc} and L_m) would be significant. The values for all three of these tests range from zero (highly significant) to one (no significance). Wilks' tables for small samples (7, pp. 263–264) were used in determining the significance of the values given in Table 3.

Instances where a lack of equivalence exists between scales of the WISC and WB in respect to the first and/or the second condition and where the hypothesis is supported are shown in Table 3 by the reference symbols *, **, and †.

Discussion

The values for the standardization groups used in the comparisons between experimental and standardization groups were taken from the age groupings as given by Wechsler (5, 6) and were the closest obtainable to the ages of the experimental groups but did not coincide exactly with these ages. In the case of the WISC, the values were from two of the standardization subgroups whose mean ages were 10½ and 13½ years respectively (6, pp. 11–13). A difference of a year therefore existed between the mean ages of the standardization groups and the mean ages of the experimental groups whose actual mean ages were 11 years, 6½ months and 14 years, 7 months. In the case of the WB, the values used were from subgroups whose mean ages were 11 and 14 years (5, pp. 122–123). The age difference is a possible but unlikely causal factor in reference to the differences obtained between experimental and standardization groups (Table 2). In addition, Wechsler included in the standardization population *S*s from the lower end of the distribution of intelligence, whereas this was not done in the case of the experimental groups. Hence, one would expect to obtain significant differences in a comparison between the two groups. On the other hand, greater variance occurs about an equal number of times in both the experimental and standardization groups which is what would be desired in a comparison of this nature.

Although this limited investigation has shown some definite results in terms of the specific criteria set forth, it is felt that before equivalence or lack of equivalence between the WISC and WB Form I within this age range can be assumed, further study should be done with the subtests of these two tests. The present investigation omitted statistical comparisons between subtests because of the lack of comparability of the WISC subtests with the WB subtests in any way other than with the use of a rank-ordering type of procedure. Raw scores are assigned weighted score values on the WB subtests prior to the application of the established age norms to the data. On the WISC, however, raw subtest scores are assigned weighted values and age norms are taken into consideration simultaneously. Thus on the WB, a set of weighted subtest scores for any one individual would be the same as the weighted scores for another individual of a different age who gave the same responses given by the first individual. The opposite holds true in the case of the WISC. Two individuals of different ages giving identical responses on all of the subtests would receive different weighted scores. Until a method is devised that permits statistical comparison of the WISC subtests with the WB subtests, it is felt that any conclusions drawn concerning equivalence of these two tests as a whole will be tenuous.

Summary

The hypothesis tested in this study stated that the WISC and WB Form I are not equivalent tests with respect to the VS, PS, and FS in the age range of 10 to 15 years. In testing this hypothesis, a sample of 40 white, American *S*s at the 11½-year age level and 40 at the 14½-year age level was used. Three criteria for equivalence were specified in which (*a*) a desired *r* between the VS, PS, and FS of the two tests was established, and the differences between (*b*) means, and (*c*) variances of the three scales were tested for significance. The hypothesis of lack of equivalence was supported to the degree the data failed to meet these three criteria. At the 11½-year age level, a lack of equivalence was established on the VS on criteria (*a*) and (*b*), and on the PS on only criterion (*b*). No lack of equivalence was found on the FS. At the 14½-year age level, lack of equivalence was established on the VS on criterion (*b*) and on the PS and FS on criterion (*a*).

Received March 24, 1955.

References

1. Delattre, Lois, & Cole, D. A comparison of the WISC and the Wechsler-Bellevue. *J. consult. Psychol.*, 1952, **16**, 228–230.
2. Graham, E. E. Wechsler-Bellevue and WISC scattergrams of unsuccessful readers. *J. consult. Psychol.*, 1952, **16**, 268–271.
3. Hollingsworth, Berneice H. An investigation of the correlation between the Wechsler Intelligence Scale for Children and the Durrell-Sullivan Reading Capacity Test. Unpublished master's thesis, Univer. of Denver, 1954.
4. Vanderhost, Leonette, Sloan, W., & Bensberg, G. J., Jr. Performance of mental defectives on the Wechsler-Bellevue and the WISC. *Amer. J. ment. Def.*, 1953, **57**, 481–483.
5. Wechsler, D. *Measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
6. Wechsler, D. *Wechsler Intelligence Scale for Children: Manual.* New York: Psychological Corp., 1949.
7. Wilks, S. S. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. math. Statist.*, 1946, **17**, 257–281.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

Diagnostic Accuracy or Diagnostic Stereotypy?

C. H. Patterson

Veterans Administration Regional Office, St. Paul, Minnesota

In the April, 1954, issue of this Journal, there appeared an article by Albert Kostlan entitled "A method for the empirical study of psychodiagnosis" (9). This study concluded that diagnostic inferences which were "better than chance" were possible using two tests and a social case history, and also using certain "minimal data" consisting of age, marital status, occupation, education, source of referral, and *the fact that the subject was a mental hygiene clinic patient* (italics mine). The original report should be consulted for details of the study.

While this study is ingenious in design, it would appear that there are two criticisms which affect the conclusions drawn.

1. The measure of chance success was taken to be 50 per cent correct in designating statements about a subject as true or false. Since significantly more than 50 per cent of two sets of statements were correctly labeled by the 20 psychologists participating in the experiment when only the "minimal data" were available, it is concluded that such data "allow inferences to be made which are more accurate than those expected on the basis of chance alone. Hence it follows that all the batteries allow inferences to be made more accurately than on the basis of chance alone, because the most errors were made under the condition of 'minimal' data" (9, pp. 86–87).

It is suggested here that the proportion correctly labeled on the basis of the knowledge that the subject was a clinic patient, which knowledge was possessed by all judges in every condition, should be used as the chance expectation. It would appear that *only if this knowledge were not available, and some of the subjects were not in fact patients,* could 50 per cent be taken as chance expectation. Inferences made on the basis of tests and the social case history should then be compared with the 77 per cent and 68 per cent correctly labeled on the basis of the "minimal data" to determine whether the results were better than chance. These comparisons were actually made by Kostlan to evaluate the relative values of the tests and case history, but are affected by the criticism mentioned below.

Kostlan states that "there was an expectation that clinicians can make inferences which are more accurate than chance simply by knowing that a person is a psychiatric patient" (9, p. 84). It is contended here that this *is* the chance expectation, rather than a pure 50 per cent. It is suggested that it is not an indication of diagnostic skill that psychologists, knowing that a subject is a clinic patient, can correctly label significantly more than 50 per cent of true-false statements taken from psychologists' reports and therapists' notes. Such statements tend to be general in nature and frequently psychological or psychiatric stereotypes or clinical clichés. The greater the number of such statements, which can be made about all clinic patients, the greater will be the percentage correctly labeled. The results would then indicate only that psychologists agree that psychiatric patients, in general, can be described by certain statements or stereotypes.

2. The conclusions that the use of tests and social case history data results in significantly better "diagnosis" than that obtained from use of only the "minimal data," and (using the therapists' statements) that using the MMPI and either the Rorschach or the Stein Sentence Completion, and the social case history, results in significantly better "diagnosis" than use of all three tests without the social case history data, are based

upon t tests of the differences between these various conditions of judging. The F test for all conditions taken together was significant.

The proper procedure for making individual comparisons following an analysis of variance has been a matter of concern to many statisticians for some time. Although Guilford (6), McNemar (11) and Lindquist (10) approve the use of the usual t test for individual comparisons following the finding of a significant F, other competent statisticians do not agree (1, 2, 3, 4, 5, 7, 13, 14, 18). It would seem that it should be clear that the ordinary t test is not appropriate where the comparisons are not random, but selected, as in the case of comparing the largest with the smallest obtained means; or in the case where all possible comparisons are made, in which case a certain proportion of the results, determined by the significance level selected, will reach significance by chance.

A number of solutions to this problem have been proposed since the first approximate, intuitive suggestion of Fisher in 1935 (5, p. 57). Fisher suggested that the number of possible comparisons be taken into consideration, so that the required probability for significance at the .05 level, for example, would not be 1 in 20, but 1 in 20 $[\frac{1}{2}k(k-1)]$ where k is the number of variables being compared. This suggestion has been developed by several writers, and is mentioned in a few texts (1, p. 239; 4, pp. 329–330; 7, p. 234), and was referred to by the present writer in an earlier note (12). The application of this procedure in the study under discussion, for example, where 10 comparisons are involved, would require that for a difference to be significant at the .05 level it would need to reach significance at the .005 level in the ordinary t tables.

Other solutions to this problem have been proposed recently. Tukey in 1949 (14) offered a solution which he later abandoned. This method has been used in several recent studies. Lindquist refers to it (10, p. 95, footnote), and Edwards in his recent text presents the method (4, pp. 330–335). It may therefore be expected that this now discarded method will persist in use for some time. In 1951 Tukey introduced another method (15), and later suggested several others, for various types of conditions (16).

Scheffé (13) has also considered this problem and introduced a method which is applicable where the interest is in all possible comparisons of pairs and combinations. He suggests Tukey's later method for use where the interest is only in the $k(k-1)$ individual comparisons. Scheffé's method is given by Jones in his chapter in the 1955 *Annual Review of Psychology* (8).

Most recently Duncan (3) has surveyed existing proposals, finding them inefficient, and has introduced his own solution. His method appears to avoid the overconservatism of most of the other solutions. It is an extremely simple method to use. However, it has been developed only for the case of equal N's in the subgroups, and needs to be extended to the case of unequal N's.

This is not the place to evaluate the merits and appropriate conditions for the use of these various methods. The writer is presently preparing a paper on this problem. It is only desired to take this opportunity to call them to the attention of psychologists having occasion to analyze individual comparisons following the F test.

In the study under discussion it would therefore appear that adequate tests of the effects of the different conditions were not made, so that the significance of the differences, and therefore the validity of the conclusions, is not known. It is not the writer's purpose to single out the author for particular censure, since the same procedure has been and is being widely used and since, at the time of the study, adequate techniques were not available, although Fisher's proposal and Tukey's 1949 method were. It is to be hoped that appropriate solutions will soon be available for all circumstances where psychologists and statisticians have been plagued by the awareness of the inapplicability of the usual t test.

Received May 6, 1955.

References

1. Anderson, R. L., & Bancroft, T. A. *Statistical theory in research.* New York: McGraw-Hill, 1952.
2. Cochran, W. G., & Cox, G. M. *Experimental designs.* New York: Wiley, 1950.

3. Duncan, D. B. Multiple range and multiple F tests. *Biometrics,* 1955, **11**, 1–42.
4. Edwards, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.
5. Fisher, R. A. *The design of experiments.* (6th Ed.) New York: Hafner, 1951.
6. Guilford, J. P. *Fundamental statistics in psychology and education.* (2nd Ed.) New York: McGraw-Hill, 1950.
7. Johnson, P. O. *Statistical methods in research.* New York: Prentice Hall, 1949.
8. Jones, L. V. Statistical theory and research design. *Annu. Rev. Psychol.,* 1955, **6**, 405–430.
9. Kostlan, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.,* 1954, **18**, 83–88.
10. Lindquist, E. L. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
11. McNemar, Q. *Psychological statistics.* (2nd Ed.) New York: Wiley, 1955.
12. Patterson, C. H. Note on "A qualification in the use of analysis of variance." *Psychol. Bull.,* 1952, **49**, 148–150.
13. Scheffé, H. A method for judging all contrasts in the analysis of variance. *Biometrika,* 1953, **40**, 87–104.
14. Tukey, J. W. Comparing individual means in the analysis of variance. *Biometrics,* 1949, **5**, 99–114.
15. Tukey, J. W. Quick and dirty methods in statistics. Part II. Simple analyses for standard designs. *Proc. Fifth Annu. Convention, Amer. Soc. for Quality Control,* 1951, 189–197.
16. Tukey, J. W. Some selected quick and easy methods of statistical analysis. *Trans. N. Y. Acad. Sci.,* Ser. II, 1953, **16**, 88–97.
17. Tukey, J. W. The problem of multiple comparisons. Unpublished notes, Princeton Univer., 1953.
18. Underwood, B. J., Duncan, C., Taylor, Janet, & Cotton, J. *Elementary statistics.* New York: Appleton-Century-Crofts, 1954.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

A Reply to Patterson

Albert Kostlan
VA Hospital, Oakland, California

C. H. Patterson (5) raises some issues, by implication, which are of importance to the research-disposed clinician. His first criticism is well taken. I differ, however, with the implication of the title, that stereotypy connotes inaccuracy. What is the diagnostic process if it is not the ordering of the patient from a supraordinate stereotype (i.e., class of events) to more subordinate ones as increments of information are accumulated and the most probable idiographic inferences emerge? Where, along this continuum, does stereotypy end? The present-day nosological categories, for all their disadvantages, are highly useful, empirically derived, stereotypes. Patterson suggests that diagnostic skill was not indicated by the condition of "minimal data." The question remains open, but it is plausible that if minimal information would permit valid stereotypy, highly accurate inferences could be extrapolated from the stereotype. The problem has relevance to the current interest in statistical versus clinical prediction (2). Statistical prediction and its logical extension, "cookbook" diagnosis (3), is a special case of stereotypy.

Patterson's second criticism is the more serious. I cannot take issue with the statistical arguments he presents, but I do feel that in a heuristic or scanning experiment (few researches in clinical psychology can be considered crucial), it is most reasonable to accept the authority of the more liberal statisticians. Of more relevance are the limitations of the experimental designs themselves, so that even "real" differences must be seen as presumptive real differences—to be studied further. What confronts us here is a situation similar to the old problem of choosing between Type I or Type II errors (4, p. 69). In the case of formal studies of our own psychodiagnostic processes, we do not have enough bathwater as it is, without running the risk of throwing away a baby—even occasionally.

In this context, it may also be appropriate to differentiate between statistical and pragmatic significance. In a recent cross-validation study of several systems for detecting organic brain pathology by means of the Rorschach, it was found that agreement between the criterion and one of the systems on a fourfold table yielded a phi coefficient of .32 (1), which was found to be highly significantly different from zero. However, statisticians will tell us that such a correlation indicates the Rorschach will do only about ten per cent better than tossing a coin. A clinician, on the other hand, is pleased to note that if he finds an "organic" record, he can be about 95 per cent certain that the patient has an organic lesion! Whither significance? In the case of the study Patterson discusses, the greatest difference that emerged (with a p of less than .01) was approximately six per cent fewer errors in favor of one battery. This is hardly of importance when the limited generalizability of the results is considered. When specific studies, notably in clinical psychology, are examined from these standpoints, the statistical objection that Patterson raises becomes academic.

Received July 29, 1955.

References

1. Fisher, J., Gonda, T. A., & Little, K. B. The Rorschach and central nervous system pathology: a cross-validation study. *Amer. J. Psychiat.*, 1955, 7, 487–492.
2. Meehl, P. E. *Clinical versus statistical prediction.* Minneapolis, Minn.: Univer. of Minnesota Press, 1954.
3. Meehl, P. E. Wanted—a good cookbook. *Amer. Psychologist,* in press.
4. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
5. Patterson, C. H. Diagnostic accuracy or diagnostic stereotypy? *J. consult. Psychol.*, 1955, 19, 483–485.

Journal of Consulting Psychology
Vol. 19, No. 6, 1955

The Manifest Anxiety and ACE Scales and College Achievement[1]

Henry E. Klugh
Alma College

and A. W. Bendig
University of Pittsburgh

Although Matarazzo, *et al.* (2) reported a significant correlation of − .25 between MAS and ACE scores ($N = 101$), Schulz and Calvin (3) found an insignificant r of .02 ($N = 98$). As additional evidence, we administered the MAS and also Gough's *Hr* scale (1) to all students enrolled in the introductory psychology course at the University of Pittsburgh during the Fall semester of 1954–55. For 184 men and women *S*s, data were available on four variables: the ACE, MAS, and *Hr* scales and the *S*s' Quality Point Averages (QPA). Raw scores on the three tests were converted to stanines on the basis of previous norm groups and Pearsonian correlations between the pairs of variables were computed. In addition, multiple correlations between the test variables and the QPA criterion were obtained.

The intertest correlations were: ACE and MAS, − .11 (not significant); ACE and *Hr*, .29 (significant at .01 level); and *Hr* and MAS, − .29 (significant at the .01 level). The coefficients of each test with the QPA criterion were: ACE, .62; *Hr*, .32 (both significant at the .01 level); and MAS, .01. Comparison of the single order and multiple correlations indicated that (*a*) a combination of the ACE and *Hr* scales is a better predictor of QPA than the ACE alone (.05 level), (*b*) adding the MAS to either the ACE or the *Hr* scales did not significantly increase the predictability of QPA, and (*c*) adding the MAS to the ACE-*Hr* combination did significantly (.05 level) increase the multiple correlation with QPA.

Our ACE-MAS correlation results fall midway between those previously reported (2, 3). A significant negative correlation between the *Hr* and MAS scales has also been found in previous samples and casts doubt upon the hypothesis that the MAS correlates negatively with timed tests and zero with untimed tests (2), since the *Hr* scale is untimed. Apparently when the ACE and *Hr* tests are combined the addition of the MAS will increase the validity of the battery by acting as a weak suppressor variable for the QPA-irrelevant variance in both the ACE and *Hr* predictors.

Brief Report
Received June 23, 1955.

References

1. Gough, H. G. The construction of a personality scale to predict scholastic achievement. *J. appl. Psychol.*, 1953, **37**, 361–366.
2. Matarazzo, J. D., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, **18**, 201–205.
3. Schulz, R. E., & Calvin, A. D. A failure to replicate the finding of a negative correlation between Manifest Anxiety and ACE scores. *J. consult. Psych.*, 1955, **19**, 223–224.

---

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 4625 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

NEW BOOKS AND TESTS

Editorial Note

After this issue, the *Journal of Consulting Psychology* will publish no more reviews of books. The task of reviewing books in all fields of psychology will be assumed by a new journal, *Contemporary Psychology, A Journal of Reviews.* Sponsored by the American Psychological Association, the new periodical will make its first appearance in 1956, under the editorship of Edwin G. Boring of Harvard University.

The *Journal of Consulting Psychology* will continue to review psychological tests, whether they are published as paper pamphlets, or as "tests in hard covers"—books which are manuals for specific tests. The plans of the journal contemplate two stages of test reviews. Brief notices, prepared mainly by the Editor, will appear promptly as in the past. Selected tests of above-average importance will later be the subjects of more extended and critical reviews, prepared by specialists on the invitation of the Editor.

The reviews which appear in the present issue were written by the Advisory Editors, who may be identified by their initials.

Books

Baldwin, Alfred L. *Behavior and development in childhood.* New York: Dryden Press, 1955. Pp. xvii + 619. $6.25.

In his Preface the author states that he attempts "to make explicit a theoretical framework which can be of help in predicting how children behave and how they develop." The writer justifies this by drawing the reader's attention to the fact that children are important, by noting that the facts of behavior do not speak for themselves, and by indicating that such facts must be seen in a theoretical context before they can lead to valid predictions. Although the author offers his work as a theoretical text, it is certainly not a text of theory. As a matter of fact, a theory of development of behavior receives no particular emphasis until chapter 22 begins on page 537; analogies, schemata, and other conceptual devices are employed in most of the earlier chapters. The book is divided into two parts. Part One has to do with how the child behaves; Part Two is concerned with prediction of personality change. The presentation of the material is systematic, well-illustrated, and in the style of an elementary textbook. Nevertheless, because of the author's rather comprehensive treatment of child behavior, the book is likely to interest the graduate student as well as the undergraduate student. Although there may be some divergence of opinion concerning the desirability of describing Baldwin's work as a theoretical text, it is probable that almost all readers will regard it as a desirable addition to the textbooks currently available in the field of child development.—*J. R. W.*

Balint, Alice. *The early years of life: A psychoanalytic study.* New York: Basic Books, Inc., 1954. Pp. vii + 149. $3.00.

Twenty years ago, the English psychoanalyst Alice Balint wrote a book (*The Psychoanalysis of the Nursery*) interpreting child behavior to parents and teachers. The present volume is a republication of this earlier work, designed to inform those who deal with children of some of the principles of child behavior and development. The book must have seemed much more novel two decades ago than it will today, for many of the principles presented as new then are well accepted now. Still, it is a straightforward, informative account of child development as the British psychoanalyst sees it. Chapter titles such as "The Oedipus Complex," "The Castration Complex," and "Identification" serve to indicate the consistently psychoanalytic viewpoint from which the book is written. Many sophisticated parents and many professional workers with children will find this volume a welcome review of psychoanalytic fundamentals.—*A. M. G.*

Burt, Sir Cyril. *The subnormal mind.* (3rd Ed.) London and New York: Oxford Univer. Press, 1955. Pp. xx + 391. $5.00.

The third edition of the now almost classic British volume by Burt has been broadened to include contemporary findings from the fields of clinical and child psychology. These welcome additions render the original title somewhat misleading, for the book covers normality, delinquency, and neurosis as well as mental deficiency and dullness. The mentalistic frame of reference within which these topics are treated makes this volume quite different from the usual, and particularly attractive to readers who have tired of a strictly behavioral approach, or who are concerned with the implications of the concept of mind for problems of educational guidance. Despite a good bit of forensic material which is unique to the English educational system, the book has much to offer to all those interested in problems of childhood.—*A. M. G.*

Bush, Robert R., & Mosteller, Frederick. *Stochastic models for learning*. New York: Wiley, 1955. Pp. 364. $9.00.

The authors consider learning to be any systematic process of behavior change which eventually reaches a condition of relative stability. This very general, descriptive concept of learning is expressed in terms of a statistical model which describes but does not seek to explain the process of behavior change. The model in effect describes a behavioral sequence by expressing the probability of a future response in a sequence as a function of the frequency of certain prior responses. The conceptual basis for this general model appears to be simple enough, and the authors' presentation is not discouragingly complicated. This discussion is particularly admirable because it provides descriptions which involve very few parameters. Since the authors go directly to their task and do not linger in obeisance before the sacred cows of learning theory, their approach is, at the very least, refreshing and may stimulate investigators to use new designs in procedure and analysis which could result in early increments to our understanding of the process of behavior change.—*J. R. W.*

Eissler, Ruth, *et al.* (Eds.). *The psychoanalytic study of the child.* Vol. IX. New York: International Universities Press, 1954. Pp. 369. $7.50.

This volume begins with two contributions from the New York Psychoanalytic Society's 1954 meeting to consider problems of infantile neurosis. An impressive group of child analysts, headed by Anna Freud, discusses this complex topic, and their recorded comments are probably the outstanding feature of this book. Of the remaining papers, mention should be made of Stevenson's study of children's treasured possessions, Sperling's presentation of a child's imaginary companion, and the Planks' paper on autobiographical accounts of emotional aspects of learning arithmetic. Both individual and group residential methods for treating disturbed children find space in the section on techniques. As contrasted with earlier volumes of the series, the 1954 book seems to contain more papers on rather specialized topics, and fewer considerations of fundamental theory of child development.—*A. M. G.*

Greenacre, Phyllis. *Swift and Carroll: A psychoanalytic study of two lives.* New York: International Universities Press, 1955. Pp. 306. $5.00.

Concerned with the lives and literary careers of the creators of Gulliver and Alice, this little book had its origins in the clinical impression of a practicing analyst that fetishistic perversions are associated with distortions of the body image. The presentation of these sensations of change in bodily size or the size of various bodily parts in the books of two enigmatic artists presented a problem in both literary criticism and psychodynamics. In dealing with this challenge, Dr. Greenacre has many suggestive and stimulating things to say about the defensive functions of the self image, the psychological determinants of satire and nonsense as forms of humor, and the relationship between impulse-anxiety and literary creativity. Like most such work, *Swift and Carroll* provides new dimensions for the understanding of books of lasting appeal and formulates new hypotheses for the comprehension of clinical complexities. Whether these hypotheses are sound or not is a question which awaits the work of investigators whose own ingenuity responds to Dr. Greenacres's.—*E. J. S.*

Heiser, Karl F. *Our backward children.* New York: Norton, 1955. Pp. 240. $3.75.

In a book planned to give parents and the interested public an over-all view of the problems that must be faced in diagnosing, training, and caring for the backward child, Heiser has tried to clear some of the confusion that parents often obtain from their contacts with specialists. He is interested in relieving parents' feelings of tension and anxiety, and giving them insight into the needs of the whole child. The book is especially recommended to those parents who are asking: "What about my slow child? What can be done for him?" There is an excellent index.—*B. M. L.*

Lorand, Sandor (Ed.). *Yearbook of psychoanalysis.* Vol. 10. New York: International Universities Press, 1955. Pp. 277. $7.50.

This tenth volume continues the tradition of the *Yearbook of Psychoanalysis* in presenting a wide variety of contemporary psychoanalytic papers. There is something here to attract any reader, although it is only the rare specialist in psychoanalytic theory who will find the majority compelling. Bernfeld's article on Freud's study of cocaine is illuminating, especially considered in conjunction with Jones's recent biographies of Freud. Notes on the Wolf Man and on Schreber add to the recorded information on these now classic cases. Of the usual series of papers on dreams, the one by Lewin seems most noteworthy for ingenuity of argument and precision of definition. Kant, Mark Twain, and W. S. Gilbert are the subjects of interesting psychoanalytic investigation. Other papers on ego function, resistance, children's early object-attachments, and various symptom manifestations complete the volume.—*A. M. G.*

Michaels, Joseph J. *Disorders of character.* Springfield, Ill.: Charles C Thomas, 1955. Pp. x + 148.

The author "concludes" that there probably is a special kind of psychosomatic disposition that permeates the delinquent individual and gives rise to specific individuation at the biological, psychological, and sociological levels. Moreover, he "concludes" that the general pattern of lack of control which characterizes persistent enuresis permeates the whole psychopathic personality. "Concludes" is a strong verb to use in sentences that are so sweepingly general, particularly on the basis of the data presented in this book. The volume, with a bibliography of 235 titles, makes reference not only to the studies on persistent delinquency but also to the pertinent theoretical literature on personality and character. The reported relationship between apprehended delinquency and persistent enuresis is shown, and is related to the phenomena of psychopathy. The author assumes a typological viewpoint, and argues that psychopathic personality and compulsive neurosis are psychobiological contrasts. The data might well lead one to *hypothesize* the existence of certain constitutional factors (among others) in delinquency and other complex behavior. The specific nature of these relationships has yet to be shown, since there are many variables that must be controlled. One important uncontrolled variable in most of the studies presented here is early interpersonal relations—a variable not emphasized by the author.—*F. McK.*

Super, Donald E. *Opportunities in psychology*. New York: Vocational Guidance Manuals, 1955. Pp. 96. $1.15.

This is one of a series of booklets designed for the intelligent choice and planning of a career. It might well serve as a model for occupational descriptions of this kind. Super presents factual material which gives the student information of wide scope in a highly readable form. It should also be informative to those in the profession. He has presented the vocation in a manner which will probably satisfy, in the main, psychologists of widely different backgrounds and interests—a difficult undertaking. He has drawn upon most of the significant publications on the activities of the profession, its prospects, rewards, fields, and organizations. The index and bibliography are brief and designed mainly for students. This volume serves a need of long duration for a booklet that may be handed to potential psychologists. In addition, it is a good contribution to the public relations of the profession.—*F. McK.*

Thorne, Frederick C. *Principles of psychological examining*. Brandon, Vt.: Journal of Clinical Psychology, 1955. Pp. v + 494. $7.50.

This textbook is designed as a complete guide to psychological diagnosis, based on the author's eclectic approach, and essentially operational in nature. Psychodiagnostic testing is relegated to a relatively minor role; observational and interview techniques are preferred and examined in unusual and useful detail. Part One covers the theory and process of psychological examining; Part Two the study of factors organizing personality integration. Each subtopic discussed is exhaustively outlined, but a final synthesis is not achieved. The author's insistence that mental health and good adjustment are more than just the absence of ill-health and defects, and require a positive integrative process, is refreshing and important.—*A. R.*

Wallin, J. E. Wallace. *Education of mentally handicapped children*. New York: Harper, 1955. Pp. xiii + 485. $4.50.

Although there is no scarcity of publications on the mentally handicapped, this book is quite special. It is written by one who has been prominently associated with the field of education of the mentally retarded from its very beginnings, more than fifty years ago. Tracing the origins of many current practices, Wallin outlines briefly, but adequately, some of the major programs for the education of mentally handicapped children, paying particular attention to legislative provisions, administrative practices, teaching methods, and curricula. This volume has particular interest for school psychologists, especially those of relatively recent vintage. Prior to the professionalization of school psychology, practitioners in "special education" frequently acquired psychological training, and many of them became the pioneers in school psychology. The modern school psychologist is well trained in clinical psychology, but often lacks close acquaintance with special education. Wallin's book compensates for much of this lack. Extensive bibliographies add considerably to the usefulness of this volume.—*M. K.*

Books Received

De Kok, Winifred. *You and your child*. New York: Philosophical Library, 1955. Pp. vii + 147. $3.75.

Farrell, B. A. (Ed.) *Experimental psychology*. New York: Philosophical Library, 1955. Pp. xi + 66. $2.75.

Sell, DeWitt E. (Ed.) *Manual of applied correctional psychology*. Columbus: Ohio Department of Mental Hygiene and Correction, no date (1955). Pp. v + 68 (paper).

Valentine, C. W. *Parents and children*. New York: Philosophical Library, 1955. Pp. xi + 212. $3.75.

Authors of Books and Tests Reviewed

Aldrich, C. K., 404
Almy, Millie, 404
Allen, Lucille, 236
Allen, R. M., 235
Allport, G. W., 318
Arbuckle, D. S., 75
Arieti, S., 318

Baldwin, A. L., 488
Balint, Alice, 488
Basowitz, H., 404
Bauman, Mary K., 154
Bellak, L., 75
Blum, Lucille H., 318
Brand, H., 75
Burrow, T., 75
Burt, C., 488
Bush, R. R., 489

Crowder, N. A., 319

Davidson, Helen H., 318
Deutsch, F., 235, 405
Donahue, Wilma T., 235
Drake, R. M., 237
Dymond, Rosalind F., 77

Edwards, A. L., 156, 235
Eidelberg, L., 76
Eisler, Ruth, 489

Fieldsteel, Nina D., 318
Frank, L. K., 76
Freeman, W., 236

Greenacre, Phyllis, 489
Grinker, R. R., 404
Gruenberg, Sidonie Matsner, 76
Guilford, J. P., 405

Harrison, R., 76
Heiser, K. F., 489
Hellersberg, Elisabeth, 76
Hertzka, A. F., 405
Holsopple, J. Q., 154
Holzinger, K. J., 319
Honzik, Marjorie P., 236

Jones, M. R., 76

King, H. E., 77
Korchin, S. J., 404

Langmuir, C. R., 78
Lewis, Hilda, 154
Lorand, S., 489
Lorr, M., 155
Lowenfeld, Margaret, 154

Macfarlane, Jean W., 236
Machover, Karen, 76
McNemar, Q., 155
Manson, M. P., 78
Maslow, A. H., 77
Meehl, P. E., 155
Meerloo, J. A. M., 236
Meyers, C. E., 237
Miale, Florence R., 154
Michaels, J. J., 489
Mosteller, F., 489
Murphy, W. F., 235, 405

Oliven, J. F., 405

Patrick, Catharine, 318
Persky, H., 404
Piaget, J., 77
Porteus, S. D., 405

Remmers, H. H., 236
Robinson, Mary F., 236
Rogers, C. R., 77
Rose, A. M., 318
Royce, J. E., 319
Rubinstein, E. A., 155

Sappenfield, B. R., 236
Savage, L. J., 237
Schafer, R., 155
Schwartz, M. S., 78
Sea, Marcella R., 237
Sheldon, J., 404
Stanton, A. H., 78
Steiner, Meta, 76
Super, D. E., 490

Taylor, W. S., 78
Thorne, F. C., 490
Thorpe, L. P., 237
Tinker, M. A., 405

Ulett, G. A., 78, 319

Wallin, J. E. W., 405, 490
Wechsler, D., 319
Wittenborn, J. R., 320